THE UNIVERSITY *of* EDINBURGH

Interpretation of Multi-Parametric Statistical Models of Large Scale Neuronal Network Activity

Joseph Cronin

Doctor of Philosophy

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

2020

Abstract

Multi-parametric statistical models have been shown to be successful in characterising the statistical properties of neuronal network activity for large scale populations, in a variety of *in vitro* and *in vivo* experimental settings. In this thesis, preliminary work explored one of the primary modelling techniques used to reproduce firing pattern distributions in binary networks, the Ising model, and outlined a novel manipulation to this analysis in the form of a distribution rescaling and renormalisation. The aim of this extension was to extract additional information about the correlation structure of typically sparse neural activity, which can often be lost due to the dominant effect that firing rate has on determining multi-neuronal firing patterns. Results here successfully uncovered a selection of simulated scenarios in which the rescaled Ising model could quantify network changes that were missed by the original model. Additionally, the interpretation of these models was also investigated through the eigendecomposition of the model's Fisher information matrix, which has been shown to offer further insights into network mechanisms through a characterisation of model parameter sensitivity structure. From here, larger scale analysis was then employed by utilising the population tracking model and the diagonal Fisher approximation, to investigate a network-wide remodelling effect in a set of multi-electrode array recordings of dissociated hippocampal cultures. Resulting entropy and divergence measures obtained from the models indicated a permanent remodelling effect upon the addition of a known chemical inhibitor, CNQX. Further to this, the use of the diagonal Fisher information approximation created an efficient framework for analysing large scale populations while maintaining sufficient information on the structure of the data, demonstrating some highly desireable qualities for use in future work. Finally, research was undertaken to characterise neural population coding impairments in a mouse model of autism spectrum disorder (ASD), by using two-photon CA^{2+} imaging of the primary visual cortex in awake behaving SynGAP$^{+/-}$ mice, in response to a number of visual stimulation protocols and an induced plasticity paradigm (monocular deprivation). Analysis techniques here involved primary response quantifications such as orientation selectivity and ocular dominance, as well as the modelling techniques presented in the earlier chapters of this thesis. While there are known physiological and behavioural differences between the two genotypes, results indicated no significant differences in the average activity, selectivity and coding capabilities of the two animal groups. However, characteristic response curves were found to exhibit different features in the SynGAP$^{+/-}$ animals,

with neurons typically responding earlier with respect to stimulus onset and displaying a lack of habituation to repeated stimulation. In summary, this study has expanded and advanced the range of analytical techniques available for the computational modelling of neural spiking activity and also contributed to refinements in sensitivity analysis techniques used for the interpretation of large scale activity of neuronal networks, as well as presenting novel results from network remodelling effects in cultured networks and visual representations in a mouse model of ASD.

Lay Summary

In the fields of neuroscience and cognition, there remains a large gap in knowledge between the physiology of small networks of neurons and the actions and behaviour of humans and animals. It therefore remains a great mystery as to how the combined activity of billions of neurons, in response to external stimuli such as light and sound, propagates through the various regions of the brain, interacting via unique neural architectures to create complex behavioural patterns. Some clue to the answer must lie in the large scale activity of neural populations within the brain, the study of which has become a central driving force in the field of neuroscience. The overarching question can be divided into two main subfields, how to obtain the experimental data from the brain and then how to analyse and interpret it. This thesis will attempt to tackle the second issue, by furthering the collection of analysis tools and models required to characterise and interpret activity in neuronal networks, using neural recordings obtained from a variety of experimental techniques. More specifically, these models will be used to monitor the structural changes in networks grown in a laboratory setting, and quantify impairments in the visual cortex of a genetically mutated mouse strain, which is a known model for an autism spectrum disorder (ASD). In doing so, the aim is not only to build on the knowledge of model interpretation in neuroscience, but to increase our understanding of how intellectual disabilities, such as ASD, are manifested in large scale neural populations. Results from this work highlighted the potential of a novel modelling technique by demonstrating its effectiveness over other methods when analysing a controlled set of simulated datasets. When analysing neuronal activity from cultured hippocampal networks, results obtained from a collection of statistical models indicated the occurrence or a permanent, network-wide remodelling effect upon the injection of a known activity inhibitor. Finally, experiments carried out on a mouse model of ASD indicated no significant differences in the overall network activity of wild type and genetically mutated mice.

Acknowledgements

First of all I would like to thank my supervisor, Dr. Matthias Hennig, for all the guidance, encouragement, ideas and inspiration throughout my PhD, it would not have been possible without him! Also, I would like to thank my examiners Dr. Arno Onken and Dr. Cian O'Donnell for their helpful critiques and input to the final work.

Many thanks to all my family and friends at home and across seas for their continued support: Mum, thanks for all the kindness and knitted hats. Dad, cheers for those rainy nights in Glasgee. Anne, thank you so much for the proof reading and inspirational talks when I really needed them! Siblings Louis, Tom, Cait and Maille, thanks for all the fun-filled visits that brightened up my time here over the years. The Crehan clan, I am very grateful to you for hosting me during my write-up months (special shout out to Cormac for being the only family member who braved Scotland with me). Grandad, thank you for first getting me interested in mathematics and science! Dr. Daniel Tanner, cheers for the laughs and the PhD advice.

A massive thank you to all the friends I made here in Edinburgh along the way: Maria Astefanoaei, Nicolas Collignon, Todor Davchev, Nathalie Dupuy, Cole Hurwitz, Pablo Leon, Ramón Mayorquin, Andrew McLeod, Jenny Sanger, Janie Sinclair, Martino Sorbaro, Akash Srivastava. I really appreciate all the coffee breaks and Teviot nights we shared over the years, you really made my time in Edinburgh worthwhile!

Finally and most importantly, thank you Kathryn, for all your support and for keeping me smiling throughout my time here.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Joseph Cronin*)

Table of Contents

Chapter 1

Introduction

1.1 Background & Motivation

Biological networks have been quantitatively studied using a variety of numerical techniques and have been investigated by researchers from numerous mathematical and scientific backgrounds. In the field of neuroscience, the 1952 series of papers by Hodgkin and Huxley [1] on the physiology of the neuron were a major breakthrough in the understanding of how the nervous system receives and transmits signals. They defined a set of non-linear differential equations that quantified the electrical properties of cells in order to describe the spatio-temporal evolution of membrane potentials through neurons. Since this initial breakthrough, a significant body of research has been focused on the development on Hodgkin and Huxley's work, particularly in relation to the behaviour of single cells and small networks of neurons. Now, however, the study of larger neural populations is generating great interest and is considered to be of fundamental importance in unravelling the underlying mechanisms that control and direct the functionality of the nervous system.

As a large proportion of mathematical models used in biology have been considered too complex to be analysed in closed form, Toni and Stumpf [2] argued that computer simulations should become the principal tool in the quantitative analysis of large-scale biological networks [3]. When considering a small number of cells, biological systems can be described by sets of equations but as the size of the network increases, statistical approaches are necessary to cope with the rapid increase in complexity that comes

with it. Early work by Hodgkin and Huxley and others suggested that modelling in computational neuroscience should be divided into two schools: one using detailed biophysical models of neuronal networks; and one in which statistical methods are used to analyse and interpret the vast amount of data accumulated from experimental neuroscience. It is now being recognised that researchers need to combine these different experimental designs into a more unified approach [4, 5]. The effectiveness of the merger and consolidation of these different techniques has been the main objective of this thesis, the analysis of which has been undertaken using both simulated and recorded neuronal network activity, with its primary focus being the further exploration into interpretation of multi-parametric statistical models of large scale neuronal activity.

The idea of the brain as a computational machine emerged in the late 20^{th} century, leading to new and exciting questions being raised, such as how networks of neurons in the brain are able to interact with each other in order to carry out computations, converting external stimuli into thoughts and actions using unique neural architectures sculpted over millions of years of evolution [6]. In recent years, a large body of work has been directed into the characterisation and understanding of the complex and diverse population dynamics exhibited in large neural ensembles, with the hope that these distinct dynamics would shed light on these processes. Researchers have since uncovered a wealth of rich dynamics within neural population activity, from diverse global couplings to criticality [7, 8, 9, 10, 11], as well as the existence of neural ensembles embedded within larger populations [12, 13, 14]. Considering neuronal firing in binary notation, Okun and others [15, 16, 17] recognised the need to capture the structure of the multi-neuron firing pattern distributions observed within neural populations by classifying the different behaviours and dependencies between neurons. Surprisingly, researchers found that a large amount of the structure in these distributions could be explained using only firing rate, pairwise correlation [18, 19] and population rate dynamics [20], avoiding the need for capturing computationally expensive higher order statistics within populations. Using only these base statistical measures, it is then possible to model entire firing pattern distributions, in a computationally inexpensive way.

The use of energy based modelling for neuronal networks draws heavily from the theories of statistical physics. Although Perretto recognised the need for a statistical mechanical approach to neural networks as early as 1984 [21], it wasn't until the early

21^{st} century that these models were studied extensively by Schneidman and collaborators who noticed the parallels between the probability distributions of firing patterns and those of so-called 'spin states' in thermodynamics. They observed, in particular, that the relation between pattern probability and spike multiplicity was identical to that of energy and entropy in the analogous thermodynamic system [22, 23, 24, 25]. Since then, a surge of publications have discussed the performance of these statistical models when modelling multi-neuronal spike trains from data recorded both *in vivo* and *in vitro* [26, 27]. The Ising model [28] was the first to be utilised due to the intuitive analogy between a spin state's dependence on external magnetic field and interaction with neighbouring atoms, and a neuron's tendency to fire within a network and its interaction (or functional connectivity) with neighbouring neurons. Although the pairwise form of the model does impose some size constraints, it remains a popular choice for capturing the multi-neuron spike distributions of neuronal networks [29, 30, 19], and inferring functional connectivity within neural populations [31], and will be one of the primary focuses of this work. The size constraints that come with the model arise due to the exponential increase in possible patterns that can be observed within a network (2^N), all of which need to be computed in order to obtain the full model. This makes it infeasible to consider networks of over ~ 20 neurons ($\sim 10^6$ patterns), hence imposing significant limitations on the information that can be obtained from them. Other limitations to the Ising model include the stationary nature of its fit, i.e. it does not account for temporal dynamics within spiking activity. While there have been attempts to circumvent these issues, by approximating the energy functions and using sampling techniques [20, 30], the need for higher order correlations becomes important for describing the statistics of larger neural population.

The issue with small scale models is that they simply cannot account for some of the large scale and complex emergent behaviours that are exhibited at the population level. In the current state of neuroscience, with the physiology and interactions of small numbers of neurons being (relatively) well understood, the study of large scale populations is arguably the most important and significant obstacle in understanding how the brain maps population-wide activity onto thoughts, actions and behaviours. This, in turn, creates the need for larger scale models, capable of characterising such activity within larger populations. One such statistical model that has recently been shown to capture the firing statistics of much larger networks ($N \sim 1000$) is the population tracking model [32]. This model leverages information from the population synchrony distribution in order to obtain a low dimensional approximation of the entire pattern

probability distribution, that would ordinarily not be computable due to the exponential increase in possible patterns for larger N. The population tracking model has been widely accepted as an accurate model of neural spiking activity and possesses a broad set of attributes that no other model contains entirely [33, 34, 35, 36] (see Gardella *et al.* for similar approach [37]). In addition, the population tracking model's low dimensional approximation of pattern distributions has been used to quantify differences in neural circuit changes in brain disorders [32] and characterise neural representation of stimuli within the mouse neocortex across development [33]. By using a subsampling of smaller, more detailed networks such as the Ising model, in collaboration with larger models of entire recorded populations, the aim is to obtain a more complete overall picture of the neural population mechanisms driving network behaviour.

In modelling any system, clear comparisons between model predictions and experimental observations should be undertaken in order to validate the model. Additionally, a vital part of any modelling lies in the ability to interpret the results and extract some further insight from the model behaviour. This aspect can sometimes become sacrificed for a higher accuracy, especially in recent times in which many so-called 'black box' methods have gained in popularity. In neuroscience, an example of this is the restricted Boltzmann machine (RBM). The RBM is another energy based model that incorporates a hidden layer into its structure, and while its ability to capture higher order statistics from larger networks is superior to the Ising model [38, 39], it is difficult to extract any significant meaning about the network mechanisms from this hidden layer [40]. This is due to the fact that is generally unclear what these hidden variables represent, or exactly what the weights connecting the visible and hidden layers are modelling, (e.g. higher order correlations, temporal dynamics, or both). The Ising model, on the other hand, contains a far more intuitive structure, with bias and interaction parameters mapping directly onto individual neuron firing rates and coincidence rates between pairs, respectively. This factor, along with a further model analysis technique (Fisher information), allows for a deeper insight into network organisation and functionality [41].

The Fisher information matrix (FIM) is a highly effective tool and offers insights through the quantification of model parameter sensitivity [42]. Originating from information theory, the FIM is a metric on probabilistic models that defines the curvature of the log-likelihood with respect to model parameters. The FIM has an extensive range of applications in science and engineering, and has recently been incorporated into

a number of breakthrough machine learning algorithms [43, 44]. In relation to neuroscience, the FIM has been used to improve decoding accuracy from the responses of macaque V1 neurons [45], as well as identifying unique, characteristic model behaviours in small groups of neurons [41, 46]. Throughout this thesis, the FIM will comprise a large proportion of the model analysis, as well as being extended and explored in a novel setting, through the use of a diagonal approximation.

Comparison of neuronal networks between animals with differing genetic mutations, or at distinct periods in time is vital for the study of learning and plasticity in the brain. The use of energy based models for comparison also has direct advantages over empirical statistical differences. For example, when comparing pattern probabilities one could simply compute a divergence measure between two empirical distributions observed at different times: the problem, however, is that many of the possible patterns are never seen and so the two distributions are difficult to compare. One of the advantages of using models is that they estimate a complete probability distribution by inferring some statistical measures from the data, despite the fact that they may never have seen certain patterns in the data before, therefore enabling a straightforward method for comparing distributions. It is worth noting, however, that while computing divergence between model probability distributions is a reasonable way to compare networks, it is difficult to distinguish whether these differences are caused primarily by firing rates or correlations. This issue is particularly prominent in neural data, which is typically sparse by nature, meaning that often the silent pattern (no neurons firing) is observed far more frequently than any other combination of spikes. This, in turn, can cause firing rate to have a dominant contribution to determining the pattern probability distribution, hence masking the effect that correlation structure has when comparing networks across time. In this thesis, a novel manipulation to the Ising model is investigated in order to uncover the role that correlation changes play in shaping pattern distributions.

Due to the billions of neurons in the human brain, it is desirable to record from as many cells as possible in order to achieve a more complete understanding of neuronal population behaviour. This has driven the need for more advanced recording techniques that can capture measurements from increasing numbers of neurons [47]. Two such recording techniques that will be utilised in this work are multi-electrode arrays and two-photon calcium imaging, which are used to obtain experimental data from hippocampal cultures and awake behaving animals, respectively. Neuronal cultures are obtained from dissociated neural tissue taken from embryonic animals and

enable a method for monitoring large networks of neurons over long periods of time in a contained laboratory environment. This separated tissue can then be grown on a suitable substrate, which enables the development of neuronal morphology and functional circuitry within the cultures. While the cultures do exhibit desired stability and plasticity at the network and single neuron levels, they do not fully capture the complexity of living brain populations due to the absence of external environmental factors and interaction with other brain regions [48, 49]. This can be supported by the famous finding from Berkes et al. [50, 51], which shows that neuronal activity in animals pre-development assimilates a sort of prior for stimulus-evoked activity, that diverges from this early form as networks become more developed and activity becomes more complex. This is essentially supporting the idea that untouched networks, in a cultured environment, will exhibit activity that is determined by pre-existing architecture, and has not yet had the opportunity to be moulded into a fully developed network through the influences of external stimuli and internal connections. Cultured networks do, however, offer a more controlled environment for experimental procedures when compared to *in vivo* recordings. In this thesis, a collection of MEA recorded datasets obtained from cultured neuronal networks are investigated in the presence and absence of a known chemical inhibitor of activity, in order to explore potential circuit-wide remodelling effects.

One particular experimental technique for recording neural activity *in vivo* that is utilised in this thesis is two-photon calcium imaging, a method that monitors levels of calcium ions in areas of the brain, through the use of genetically encoded fluorescent molecules. During such experiments, hundreds of individual, identifiable neurons can be observed within a single field of view and recorded while animals undergo a variety of perceptual tasks [47, 52]. This allows individual neurons to be tracked over long periods of time in response to induced plasticity mechanisms [53]. In this work, population coding and plasticity deficits are investigated in a mouse model of a monogenic autism spectrum disorder, by monitoring the activity of awake behaving SynGAP$^{+/-}$ mice, while they undergo a series of visual stimulation experiments and a monocular deprivation paradigm.

1.2 Outline

The aim of this thesis is to investigate the use of parametric statistical models applied to neuronal network spiking activity and more specifically, how these models can be interpreted in order to uncover additional information about underlying neural circuit mechanisms. Chapter 2 introduces the main models utilised in this work, the Ising model and the population tracking model, as well as some further analysis techniques, such as Fisher information, eigendecomposition and Kullback-Leibler divergence. Additionally, chapter 2 outlines the primary *in vivo* and *in vitro* recording procedures used to acquire the neural datasets that are to be modelled.

In chapter 3, a novel modification to the widely used Ising model is presented and tested on a controlled simulated dataset. The model is then used to quantify changes in correlation structure in dissociated hippocampal cultures in response to induced increase in inhibition. Chapter 4 then extends this analysis on cultured networks to incorporate larger scale activity by utilising the population tracking model and the diagonal Fisher information approximation, both of which are capable of analysing up to 500 neurons without an unrealistic increase in computational cost.

Finally, chapter 5 focuses on the characterisation of coding and plasticity deficits in a mouse model of a monogenic autism spectrum disorder, SYNGAP1 haploinsufficiency, by using two-photon Ca^{2+} imaging of the primary visual cortex of awake behaving mice in response to visual stimuli and an induced plasticity mechanism.

Chapter 2

Statistical Modelling of Neuronal Spiking Activity

2.1 Probabilistic Framework for Spiking Activity

Multi-parametric statistics encompasses an extensive range of models whose purpose is to capture dependencies between variables. In relation to spike trains it is first necessary to construct a probabilistic framework in order to describe the statistics and correlations of a group of spiking neurons. For a network containing N observable nodes (neurons), the network state vector is defined as $\mathbf{x} \in [-1,1]^N$. That is, the current state of the network can be represented by a binary array indicating whether each neuron is firing inside a given time bin $\delta\tau$ ('1' indicates firing, '-1' indicates not firing). Using this representation the predicted probability of each state vector $P_{\text{Model}}(\mathbf{x})$ can be estimated for all possible patterns within a given network, as an approximation of the true distribution $P_{\text{True}}(\mathbf{x})$, using information from the empirical distribution $P_{\text{Emp}}(\mathbf{x})$, obtained from a histogramming of the observed patterns within the dataset. There are a number of ways in which this can be achieved, two of which will be the main focus of this work. The first model, called the Ising model, requires the estimation of firing probabilities for each neuron i, $r_i = E[x_i]$, and the joint firing probabilities for all pairs of neurons i and j, $c_{ij} = E[x_i x_j]$, in order to define a distribution over all possible patterns. This model has been shown to accurately predict pattern probabilities for small groups of neurons, even accounting for higher correlations within the

data [29, 19]. The second model, the population tracking model, again estimates the firing probabilities but additionally leverages information from the network synchrony distribution $p(k)$, where k is the number of coactive neurons $k = \sum_{i=1}^{N} x_i$, and the conditional synchrony distribution $p(x_i|k)$, in order to obtain an approximation for much larger networks [32].

As a first step, neural membrane voltage signals are converted into binary events via thresholding or some more sophisticated techniques. The idea is then to represent spikes as a binary matrix of spike trains of size $[N \times t/\delta\tau]$ by binning the spikes of N neurons into time bins of size $\delta\tau$ over a time period of t seconds. The choice of $\delta\tau$ should be such that it is big enough to capture synaptic interactions but small enough to avoid too many polysynaptic events, ($\delta\tau \approx$10-20ms). Hence, these spike trains are readily convertible into the training matrices required for statistical models.

In the following chapter, a concise mathematical definition of the principal models utilised in this work will be presented, as well as some of the techniques used for their analysis and interpretation. Additionally, the core methods for acquiring the *in vivo* and *in vitro* datasets that will be analysed are outlined.

2.2 Energy Based Models

Energy based models encode dependencies between variables through the definition of an energy function $E(\mathbf{x})$ over all possible input units. Using this energy function a probability distribution can be defined over inputs, inspired by the Boltzmann distribution from statistical physics for a closed system of average energy E: $P \propto exp(-E/k_BT)$, where T is the temperature and k_B is the Boltzmann constant [54]. The equivalent quantity for energy based models drops these associated physical dimensions and has $k_BT = 1$, resulting in the energy function:

$$P(\mathbf{x}) = \frac{1}{Z} e^{-E(\mathbf{x})} \tag{2.1}$$

where Z is called the partition function.

$$Z = \sum_{\mathbf{x}} e^{-E(\mathbf{x})} \tag{2.2}$$

From here the model undergoes a training process that shapes the energy function by altering its parameters in order to obtain a desirable probability distribution. For example this could be the energy function that maximises the log likelihood of the probability, or in some cases the entropy.

2.3 Ising Model

The function $E(\mathbf{x})$ can take on a number of different forms corresponding to different network structures, one such form is the Ising model, a physical model used to describe ferromagnetism by calculating the spin states of atoms located in an approximated two dimensional lattice. The model takes into account the sum of spins over all atoms plus the sum of the product between connected neighbour interactions for each atom, multiplied by some physical constants [55]. In terms of neurons in a network, this energy function is given by

$$E(\mathbf{x}) = -\sum_i h_i x_i - \sum_{i,j} x_i J_{ij} x_j \tag{2.3}$$

where $h_i =$ set of biases for each neuron and $J_{ij} =$ set of interaction weights between all neurons. Conceptually, these parameters can be interpreted as a neuron's excitability or tendency to fire (bias terms h_i), and a neuron's functional connectivity to other neurons in the network (interaction terms J_{ij}). With this form of energy function, given a set of training data, the goal is then to optimise some objective function by adjusting parameters accordingly. In this work, models were trained using a maximum entropy fitting procedure, where the objective is to maximise the entropy of a distribution given a set of statistical constraints. The entropy of a distribution is defined as

$$H = -\sum_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) \tag{2.4}$$

and can be interpreted as a measure of uniformity or lack of structure within a distribution. In other words, the idea is to match measurable statistical quantities between the model and data but otherwise keep the model distribution as unstructured as possible (equivalent to maximising entropy). For the Ising model, these constraints are the first and second order statistics of the observed training data:

$$\langle x_i \rangle_{\text{data}} = \frac{1}{T} \sum_{t=1}^{T} x_i^t \tag{2.5}$$

$$\langle x_i x_j \rangle_{\text{data}} = \frac{1}{T} \sum_{t=1}^{T} x_i^t x_j^t \tag{2.6}$$

An alternative (but equivalent) formulation can be derived using Lagrange multipliers λ_i and λ_{ij} to maximise the entropy with constraints, leading to a distribution with an identical form to the Boltzmann distribution where the Lagrange multipliers are equivalent to the field and interaction parameters h_i and J_{ij} [56]. The model was fitted by updating parameters according to the iterative scaling algorithm (below) [57], and recalculating the model observables $\langle x_i \rangle_{\text{model}}$, $\langle x_i x_j \rangle_{\text{model}}$ iteratively until a suitable match between model and data statistics was obtained. The tolerance value was set so that the relative difference between model and data marginals was $< 10^{-5}$, this was to ensure the model reproduced spike and coincident rates to within a precision of ~ 1 spike [41].

$$\delta h_i = \eta \times \text{sign}\left(\langle x_i \rangle_{\text{data}}\right) \times \log\left(\frac{|\langle x_i \rangle_{\text{data}}|}{|\langle x_i \rangle_{\text{model}}|}\right) \tag{2.7}$$

$$\delta J_{ij} = \eta \times \text{sign}\left(\langle x_i x_j \rangle_{\text{data}}\right) \times \log\left(\frac{|\langle x_i x_j \rangle_{\text{data}}|}{|\langle x_i x_j \rangle_{\text{model}}|}\right) \tag{2.8}$$

Here, the learning rate η was set to be < 1 to prevent the algorithm from becoming unstable and adjusted in order to achieve the desired tolerance within a set number of iterations (50,000).

2.4 Model Interpretation

One of the benefits of the Ising model is that it defines a probability distribution over all 2^N possible patterns that could be generated by a network, even if these patterns were not observed in the training data which, due to the sparsity of neural data, is the case for many potential firing configurations. This enables a direct method for comparison between networks at different points in time and offers an efficient solution to the case in which restrictive datasets prohibit the exact computation of probability distributions. Additionally, the analytical form of the model allows for further insightful analysis from which hypotheses can be drawn about circuit mechanisms underlying network function. This section outlines some of the aforementioned techniques for model comparison and analysis, namely Fisher information and Kullback-Leibler divergence.

2.4.1 Fisher Information Matrix

For probabilistic models, such as the energy based models described previously, the Fisher information matrix (FIM) is an important tool for assessing model fit and parameter sensitivity analysis. A probabilistic model defines a probability over observables $\mathbf{x}$ with a parametric form $p(\mathbf{x}|\Theta)$, where Θ is a set of parameters. $p(\mathbf{x}|\Theta)$ is often referred to as the likelihood, detailing how probable a given set of observations is for parameter values Θ. These parametric models form a Riemannian manifold and therefore have a coordinate system (defined by the parameters) and a distance metric [58]. This distance metric is given by the FIM [59], which has the form

$$f_{lm}(\Theta) = -\sum_{\mathbf{x}} p(\mathbf{x}|\Theta) \frac{\partial \log p(\mathbf{x}|\Theta)}{\partial \theta_l} \frac{\partial \log p(\mathbf{x}|\Theta)}{\partial \theta_m} \tag{2.9}$$

In terms of information geometry, this quantity can be thought of as the distance (or distinguishability) between the model and a nearby model in which the parameters have undergone an infinitesimal change [59, 60]. In other words, this expression gives the curvature of the log-likelihood of the model with respect to its parameters, hence giving a measure of the model's sensitivity to changes in these parameters.

For the case of the Ising model the parameters $\Theta = \{h_i, J_{ij}\}$, hence, for a model of N neurons the FIM will have the following form:

$$F = \begin{pmatrix} f_{h_1,h_1} & \cdots & f_{h_1,h_N} & f_{h_1,J_{1,2}} & \cdots & f_{h_1,J_{N-1,N}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{h_N,h_1} & \cdots & f_{h_N,h_N} & f_{h_N,J_{1,2}} & \cdots & f_{h_N,J_{N-1,N}} \\ f_{J_{1,2},h_1} & \cdots & f_{J_{1,2},h_N} & f_{J_{1,2},J_{1,2}} & \cdots & f_{J_{1,2},J_{N-1,N}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{J_{N-1,N},h_1} & \cdots & f_{J_{N-1,n},h_N} & f_{J_{N-1,N},J_{1,2}} & \cdots & f_{J_{N-1,N},J_{N-1,N}} \end{pmatrix} \tag{2.10}$$

where F is an $N_F \times N_F$ symmetric matrix and $N_F = N(N+1)/2$ is the total number of parameters for an Ising model with N units. These quantities can be shown to be equal to the covariance matrix of the random variables associated with each of the parameters and can be calculated directly from the model [61]:

$$f_{lm}(\Theta) = < X_l X_m >_{model} - < X_l >_{model} < X_m >_{model} \tag{2.11}$$

where $X_l = x_i$ for the parameters h_i and $X_m = x_i x_j$ for the pairwise parameters J_{ij}. The FIM can also be found approximately by calculating the expectations over the measured probability distributions. In order to visualise this structure, Fig. 2.1 shows an example FIM computed from an Ising model fit to a network of $N = 10$ neurons obtained from cultured hippocampal recordings.

From here, further analysis of the FIM via an eigendecomposition can uncover additional structure within the model, which will be demonstrated in the following section.

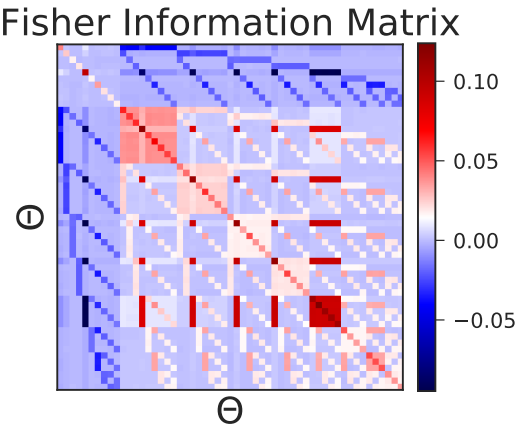


Figure 2.1: Example FIM computed from Ising model fit to cultured hippocampal data

2.4.2 Fisher Information Matrix Eigendecomposition

Following from the definition of the FIM, it is possible to examine the shape of the local parameter space, via an eigendecomposition. This property is highly informative as it can uncover whether the parameter space is uniform, suggesting the model is equally sensitive to changes in all parameters, or heterogeneous. In the latter case, in which the sensitivity to changes in parameters is highly non-uniform, an eigendecomposition can also give an ordering of importance, identifying which parameter combinations give the highest sensitivity, i.e. the strongest influence on the model behaviour. An eigendecomposition of a diagonisable matrix (such as the FIM) is a factorisation of the matrix into a set of orthogonal vectors $\mathbf{v}$ and scalars λ such that the following equation holds

$$F\mathbf{v} = \lambda\mathbf{v} \tag{2.12}$$

Here $\mathbf{v}$ and λ are said to be the eigenvectors and eigenvalues of F. In other words, there exists a set of eigenvectors $\mathbf{v}$ that, under the linear transformation F, are simply a scalar multiple of themselves, only changing in magnitude, not direction. In terms of the FIM, these eigenvectors correspond to directions along the manifold in parameter space and the eigenvalues represent the magnitude of change along these vectors. This is an essential aspect of the eigendecomposition as, in a similar manner to principal component analysis, these eigenvectors define dimensions in parameter space that give the dominant contributions to the FIM. The eigenvalues then give the ordering, or importance of these dimensions, and can be highly informative about the model structure. Fig. 2.2 shows the first eigenvector of the FIM seen in Fig. 2.1, note here that the eigenvector has been reshaped from $[1 \times N_F]$ to $[N \times N]$ so that the eigenvectors map intuitively into a form resembling the bias and pairwise interactions of each neuron.

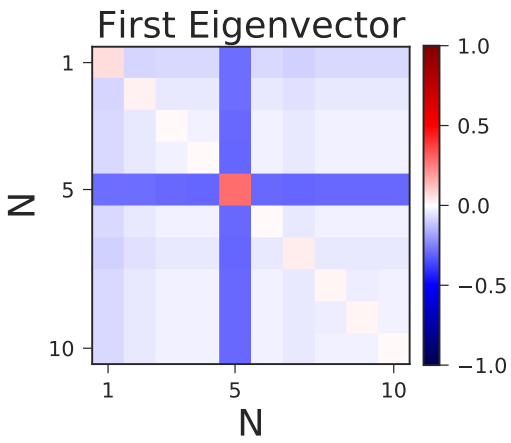


Figure 2.2: First eigenvector of the FIM computed from an Ising model fit to cultured hippocampal data

2.4.2.1 Sloppiness in Neuronal Networks

A more general framework similar to the one presented above looks at the cost function that measures a model's fit to observed data, for example a sum of squares

$$C(\Theta) = \sum_i (\hat{y}_i(\Theta, t_i) - y_i)^2 \tag{2.13}$$

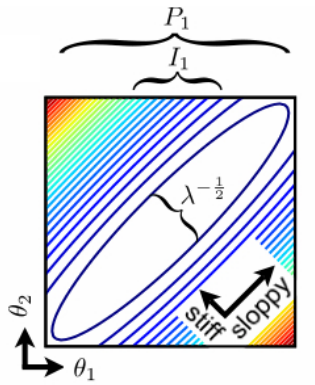


Figure 2.3: Contour plot of cost function. Stiff direction crosses many contours with only a small movement, behaviour changes very little along a sloppy direction (few contours crossed) [62].

where $\hat{y}_i(\Theta, t_i)$ is the model prediction of the observed data point y_i. This measure can also take other forms, more suitable to evaluating neuronal networks, such as spike distance measures or distance formulas relating other observable quantities like rates and correlations. The Hessian matrix of this cost function can then be calculated using the second order partial derivatives with respect to parameters θ_i and θ_j, $\frac{\partial^2 C}{\partial \theta_i \theta_j}$. Fig. 2.3 shows an example contour plot of a cost function near to the best fit for a so called 'sloppy' model, where contour surfaces are described by the eigenvectors and eigenvalues of the Hessian. The ellipsoid form of these contours are typical of sloppy models, stretching out far along 'sloppy' (highly insensitive) directions and narrow in 'stiff' (highly sensitive) directions. In terms of model behaviour, this corresponds to a large change when moving along stiff directions (crossing many contour lines) and a very small change when moving along sloppy directions [63, 64, 65].

Sloppiness in a network model can then be identified through the characteristic span of its FIM eigenvalues, with the stiff parameters corresponding to the largest eigenvalues and the sloppy parameters given by lower values. Fig. 2.4 shows the average FIM eigenvalue rank for a number of model fits sampled from the same hippocampal recordings as discussed above (note that above only one group is examined). Previous work on this topic has identified such characteristics in models of spontaneously active neuronal networks [41], from which insightful conclusions can be drawn about network stability and function. In a similar manner, the FIM is examined throughout this thesis and interpreted in order to characterise network behaviour in a number of different scenarios.

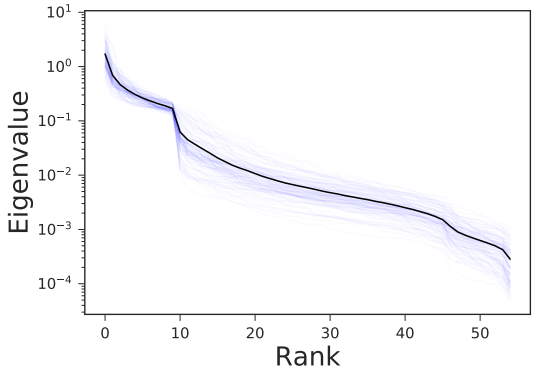


Figure 2.4: Ranked eigenvalues of a number of FIMs computed from Ising model fits to cultured hippocampal data. Average shown in black

2.4.3 Diagonal Approximation

As stated earlier, the FIM for the Ising model can be show to equal the covariance matrix of the random variables associated with each of the parameters, and can be approximated quite accurately from the data. Taking this approximation it is then possible to extend the size of the models considered to larger N and compute the FIM directly from spiking activity. However, this method still imposes size constraints due to the $N_F(N_F+1)/2$ entries of the FIM that need to be computed (note the FIM is symmetric). To avoid this, a simple approximation is to only consider the diagonal elements for the FIM, reducing the number of computations to $N_F = N(N+1)/2$. The resulting FIM will have the form:

$$F' = \begin{pmatrix} f_{h_1,h_1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & f_{h_N,h_N} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & f_{J_{1,2},J_{1,2}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & f_{J_{N-1,N},J_{N-1,N}} \end{pmatrix} \tag{2.14}$$

Clearly, only terms that include the 2^{nd} derivatives with respect to each parameter are considered. The resulting diagonal approximation is extremely efficient to compute (scales with N^2 as opposed to N^4 for the full FIM) but is only valid if sufficient data is available. In terms of geometry, the diagonal of the FIM is the intersection between the ellipse determined by the full FIM and the parameter axes, and so is not always

accurate. However, depending on the parameter space the diagonal has been shown to maintain a satisfactory level of information about the structure of the data, performing comparably well in tasks that use the FIM to compute natural gradients [66, 67] or using elastic weight consolidation in sequential learning tasks [44].

2.4.4 Fisher Overlap

Fisher overlap is a similarity measure utilised in order to assess the closeness of two Fisher matrices corresponding to two different models, hence detailing the likeness of their parameter sensitivity structure. The measure is obtained by first normalising the two Fisher matrices in question, F_1 and F_2, to have unit trace and computing the Fréchet distance between the two resulting matrices ($\hat{F}_1$ and $\hat{F}_2$) [68], as follows:

$$d^2(\hat{F}_1, \hat{F}_2) = \frac{1}{2}\mathrm{tr}\left(\hat{F}_1 + \hat{F}_2 - 2(\hat{F}_1\hat{F}_2)^{1/2}\right) \tag{2.15}$$

$$d^2(\hat{F}_1, \hat{F}_2) = \frac{1}{2}\|\hat{F}_1^{1/2} - \hat{F}_2^{1/2}\|_F \tag{2.16}$$

The Fréchet distance is a metric on the space of positive-semidefinite matrices and is bounded by 0 and 1. The Fisher overlap is then defined as $1 - d^2$ and defines a measure of similarity between two model parameter sensitivities, with a value of 0 indicating the two sets of parameters are not overlapping and a value of 1 signalling that $F_1 = \alpha F_2$ where $\alpha > 0$ [44].

2.4.5 Kullback-Leibler Divergence

The Kullback-Leibler divergence is a distance metric on probability distributions, used to measure how distinguishable two probability distributions are from each other, and is defined as

$$D_{KL}(P_1||P_2) = \sum_{x \in X} P_1(x) \log\left(\frac{P_1(x)}{P_2(x)}\right) \tag{2.17}$$

where P_1, P_2 are two discrete probability distributions over X. KL divergence is used widely within neuroscience and is extremely useful in quantifying differences between networks in different conditions, for example a population of *in vivo* V1 neurons

before and after an induced plasticity protocol or a set of cultured neurons undergoing different chemical treatments. KL divergence is always non-negative, $D_{KL}(P_1||P_2) \geq 0$, but is not a regular distance measure due to the fact that it does not necessarily satisfy $D_{KL}(P_1||P_2) = D_{KL}(P_2||P_1)$ or the triangle inequality $D_{KL}(P_1||P_2) + D_{KL}(P_2||P_3) \geq D_{KL}(P_1||P_3)$. However, the KL divergence becomes symmetric for two nearby models P_Θ and $P_{\Theta+\delta\Theta}$ where it also can be shown to approximate the FIM [60]. Another divergence measure that is commonly used is the Jensen-Shannon divergence, given by

$$D_{JS}(P_1||P_2) = \frac{1}{2}\sum_{x\in X} P_1(x)\log\left(\frac{P_1(x)}{P_M(x)}\right) + \frac{1}{2}\sum_{x\in X} P_2(x)\log\left(\frac{P_2(x)}{P_M(x)}\right) \tag{2.18}$$

The above divergence is based on the average KL divergence between two distributions and their average distribution $P_M = \frac{1}{2}(P_1 + P_2)$ and has the advantage of being symmetric and bounded between 0 and 1 [69].

2.5 Population Tracking Model

One of the main disadvantages of the pairwise model is the size constraint that arises from computing 2^N exact pattern probabilities. The population tracking model [32] is a statistical model that is highly advantageous for neural population data due to its ability to capture neuronal network distributions of up to 1000 neurons, using a significantly reduced number of computations (N^2 compared to 2^N). The model contains 2 sets of parameters: N parameters to fit the synchrony distribution $p(k)$ that represents the probability that k neurons are simultaneously active, and $(N^2 - N)$ parameters to describe the conditional probabilities that any neuron is active given the total number of active neurons k. This characterises a neuron's dependency on the population rate, identifying whether neurons act more in unison ('chorist') or independently ('soloist') from the overall population [15]. The full model is then defined by the following equation for any given pattern $\mathbf{x}$

$$P(\mathbf{x}) = \frac{p(k)}{a_k}\left(\prod_{i=1}^{N} p(x_i|k)^{x_i}[1 - p(x_i|k)]^{1-x_i}\right) \tag{2.19}$$

where $k = \sum_{i=1}^{N} x_i$ and a_k is a normalising constant defined as the sum of the probabilities of all $\binom{N}{k}$ patterns in the set $S(k)$, where $\sum_{i=1}^{N} x_i = k$ and neurons are assumed conditionally independent:

$$a_k = \sum_{\mathbf{x} \in S(k)} \left(\prod_{i=1}^{N} p(x_i|k)^{x_i} [1 - p(x_i|k)]^{1-x_i} \right) \tag{2.20}$$

Given a spike train matrix of N neurons across T time bins (note here the $[-1, 1]$ notation becomes $[0, 1]$), the model also has a quick and simple fitting procedure. For the N parameters characterising the synchrony distribution of the data, the population rate of the network is converted into a histogram with each value of k given a count c_k. These counts are then normalised by the total number of time bins giving

$$\hat{p}(k) = \frac{c_k}{T} \tag{2.21}$$

However this formulation can be problematic as neural data is often sparse and so for larger values of k the probability becomes very small. Due to the finite sample size of the data it is likely that some of these patterns will not be observed so to avoid assigning zero probabilities a Dirichlet prior distribution is assumed for $\hat{p}(k)$. The posterior mean estimate for each parameter is then

$$\hat{p}(k, \alpha) = \frac{c_k + \alpha}{T + N\alpha} \tag{2.22}$$

where α is a small positive constant (set as $\alpha = 0.01$). Similarly, for the $(N^2 - N)$ parameters that determine the conditional probabilities $p(x_i|k)$ the simplest method to fit the data is by histogramming. For each value of k (1 to N) the patterns with k active neurons are isolated and then the number of times each neuron is active within this subset of patterns is counted (d_k).

$$\hat{p}(x_i|k) = \frac{d_{i,k}}{T_k} \tag{2.23}$$

where T_k is the number of time bins within where k neurons are active. To avoid 0 probabilities given to rare events, regularisation was again used in the form of a beta prior distribution over each $p(x_i|k)$, giving a posterior mean estimate for the parameters:

$$\hat{p}(x_i|k,\beta_0,\beta_1) = \frac{d_{i,k}+\beta_1}{\beta_0+\beta_1+T_k} \tag{2.24}$$

where the two hyperparameters β_0 and β_1 are calculated by assuming a prior mean and variance of $\mu = k/N$ and $\sigma^2 = 0.5\mu(1-\mu)$, giving $\beta_0 = \frac{\mu}{\sigma^2}(\mu-\mu^2-\sigma^2)$ and $\beta_1 = \beta_0(\frac{1}{\mu}-1)$.

2.5.1 Entropy & Divergence

Another advantage of the population tracking model is that it allows for the estimation of measures over the entire spike pattern probability distribution, such as entropy and Kullback-Liebler divergence. The Shannon entropy over a distribution is defined as:

$$H = -\sum_{i=1}^{2^N} p(x_i)\log_2 p(x_i) \tag{2.25}$$

and is a measure of the uniformity of a probability distribution (maximum entropy distribution will have all patterns equally probable, i.e. the uniform distribution). The entropy can also be interpreted as the upper limit on the amount of information a network can encode [70]. Due to the enormous amount of possible patterns for networks of large N, the above expression cannot be computed directly. However, due to the form of this model the distribution of patterns will tend towards the sum of a set of log-normal distributions, one for each value of k. This approximation is accurate for most values of k, but may break down for values of k close to 0 and N, which is problematic due to the sparsity of neural data (low values of k are common). For low values of k (e.g. $k \leq 3$), distributions can be calculated by brute force enumeration, for higher values (e.g. $3 < k \leq 10$) they can be estimated using a separate empirical fit with a low dimensional parametric model, e.g. mixture of Gaussians (note this is approximate for discrete data). This is to approximate the multi-modal or skewed distributions observed at these values of k. Using this approximated distribution it is now possible to estimate the population entropy of the network, which can be broken down as follows

$$H = H_k + H(p(\mathbf{x}|K)) \tag{2.26}$$

where H_k is the entropy of the synchrony distribution, given by

$$H_k = -\sum_{k=0}^{N} p(k)\log_2 p(k) \tag{2.27}$$

and $H(p(\mathbf{x}|K))$ is the conditional entropy of the pattern distribution given k

$$H(p(\mathbf{x}|K)) = -\sum_{k=0}^{N} p(k)H(p(\mathbf{x}|K=k)) \tag{2.28}$$

Here H_k is easily computable and $H(p(\mathbf{x}|K))$ can be obtained via numerical integration. This allows for a low dimension approximation of the population entropy for a network of up to 1000 neurons.

As stated previously, Kullback Leibler divergence is common metric used to compare probability distributions and is of particular interest when comparing network spiking activity across time (during monocular deprivation) and during different stimulation protocols (right eye vs. left eye stimulation, spontaneous vs. evoked activity). The population tracking model allows for an approximation of this quantity between two model distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ by breaking up the divergence into individual components for each value of k:

$$D_{KL}(p||q) = \sum_{k=0}^{N} \tilde{D}_{KL}(p||q)_k \tag{2.29}$$

Here, each value $\tilde{D}_{KL}(p||q)_k$ can be obtained through sampling.

2.6 Neuronal Network Recording Techniques

An enormous amount of neuroscience research is focused on developing novel recording techniques in order to unlock the wealth of data that lies within the brain. These techniques are divided into two major types: *in vivo*, the study of living organisms such as awake behaving animals and *in vitro*, the study of cultured cells outside of a living organism. The two main recording techniques utilised in this thesis are discussed in the following sections.

2.6.1 Multi-electrode Array Recordings

Multi-electrode arrays (MEAs) are a type of device frequently used for the recording of neural signals both *in vivo* and *in vitro*. These devices consist of 100s - 1000s of microelectrodes, capable of capturing extracellular changes in voltage due to the flow of current in and out of cells [71, 72].

For the datasets acquired in this thesis, recordings were carried out on the Bio-Cam4096 platform using Active Pixel Sensor (APS) micro-electrode array chips, type BioChip 4096S (3Brain GmbH, Switzerland). These chips use 4096 micro-electrodes of size $21\mu m \times 21\mu m$ arranged on a 64×64 grid, with channel separation $42\mu m$ on a $2.67mm \times 2.76mm$ active area. The chips record with a $7.022kHz$ sampling rate with experiments typically lasting 15-20 minutes per session. Voltage signals at each channel were converted into spike events via spike detection algorithms developed by Muthmann *et al.* [73], with the time and channel of each event being recorded and converted into binary spike matrices via a binning procedure. The following sections will detail experiments carried out using MEAs to record neuronal network activity *in vitro* from cultured hippocampal cells.

2.6.2 Neuronal Cultures

A widely used method for studying living neurons involves the dissociation of neural tissue from embryonic animals. This separated tissue can then be grown on a suitable substrate, which enables the development of neuronal morphology and functional circuitry within the cultures. Such *in vitro* experimental methods allow for the monitoring of hundreds of neurons over large time periods (days - weeks), while the cultures exhibit desired stability and plasticity at the network and single neuron levels [74, 75]. (Note that there are differences between circuits grown in cultures and those found in living animals [48, 49])

Neurons in hippocampal cultures typically grow axons and dendrites at around 2 days in vivo (DIV) and exhibit spiking activity at around 3 - 7 DIV [76]. From here, networks exhibit irregular population bursting which is then followed by more periodic activity [77, 78]. For the experiments examined in this thesis, neural cultures were prepared using dissociated embryonic rat brain tissue at embryonic day 18. Cells were plated on APS MEAs prepared with polyethyleneimine (PEI) as an adhesion factor

and samples were then incubated in Neurobasal medium with 1% Glutamax 2% B-27 medium, in a humidified atmosphere with 5% CO_2 and 95% O^2 at 37$_\circ$C. Every 6 days, 50% of the medium was changed to aid the development of a healthy culture.

For the main experiment, an initial baseline recording was obtained (at 0h) for each sample and subsequent recordings were carried out at increasing time points. 5 μM of CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) was applied to a subset of the cultures at one or more time points over the duration of the experiment and washed out by twice replacing two thirds of the culture's medium. The time points of CNQX addition and washouts are detailed in Table 2.1. The initial experiment (chip 77) was carried out by Alessandro Maccione, Dagmara Panas and Oliver Muthmann while the remaining experiments were performed by Hayder Amin and Alessandro Maccione in the laboratory of Luca Berdondini (Italian Institute of Technology Genova).

Chip	Age	Time (m)	Time after first rec (h)	Time of CNQX (h)
77	25	20	0 28 30 34 72 77	4 - 28, 72 - 96
136	21	15	0 2 20 44 50 68 92	-
140	21	15	0 2 20 44 50 68 92	0 - 48
184	?	15	0 48 144 312 480	-

Table 2.1: Experimental protocols for culture recordings. Here, red entries indicate recordings during CNQX application and blue entries indicate first recording after a wash. CNQX was applied with a concentration of 5μM, except for the second CNQX application for Chip 77 which was 10μM. The age of the cultures refers to the start of the experiment

CNQX is a competitive AMPA/kainate receptor antagonist [79, 80] that blocks excitatory synaptic transmission, leading to a decrease in average activity as well as in the number of neurons being active [81, 82, 83].

2.6.3 Two Photon Calcium Imaging

Despite the excellent temporal and spatial resolution of membrane voltage recording devices they also come with some disadvantages. Firstly, the intrusive nature of the devices in *in vivo* experiments makes it difficult to monitor animals over long periods

of time and furthermore, these techniques are essentially blind to neuron locations, relying on highly complex spike sorting algorithms to identify which voltage traces belong to which neurons. In order to achieve an accurate recording of individually identified neurons it is necessary to utilise neural imaging techniques, such as two-photon imaging of genetically encoded calcium indicators. Using such techniques allows for the recording of large scale neuronal population activity with single cell resolution [47, 52], enabling individual neural responses to be monitored over time in response to induced plasticity mechanisms [53]. In order to do this the fluorescence of genetically encoded calcium indicators is measured using two-photon microscopy, with changes in fluorescence corresponding to changes in intracellular free calcium levels, which is used as an approximate readout of neuronal activity [84]. Of course, these benefits also come at a cost of temporal resolution, with imaging frame rates far lower than the sampling rate of MEAs. Additionally, the complex nature of calcium indicator dynamics, coupled with the noise associated with the imaging process makes it difficult to accurately infer underlying spiking activity. Techniques on how this is achieved will be discussed later in this thesis.

Chapter 3

Rescaled Energy Based Models

3.1 Introduction

In sparse neural data the silent pattern (all -1's or all 0's depending on notation) is observed significantly more than any other pattern and hence dominates the network's multi-neuron firing distribution. This property of neural data causes difficulty when comparing two firing distributions across time, for example, as differences in distributions are in turn largely dominated by differences in the silent pattern probability. This effect can mask contributions due to a change in underlying correlation structure, which is of great interest when studying the rewiring of networks due to perceptual learning, or in response to induced inhibition in cultured networks. The proposed solution is to reduce this dominance by normalising out the field interactions (h_i's) from the Ising model in order to obtain a clearer picture of differences in probability distributions arising from changes in correlation structure, that would otherwise be masked in a change in firing rate.

In this chapter, we explore the potential of such a renormalisation by scaling the Ising model pattern distribution with that of a simple independent model, in order to extract the captured correlation structure of spiking data. Preliminary analysis of such models fit to both simulated dichotomised Gaussian networks and recorded hippocampal cultures is presented and discussed, with results highlighting the interesting advantages of this novel model manipulation.

3.2 Methods

3.2.1 Ising Model Renormalisation

Continuing from the formulation of the Ising model energy $E_{pair}(\mathbf{x})$ found in the previous chapter, the energy function of the independent model $E_I(\mathbf{x})$ is defined as follows.

$$E_{pair}(\mathbf{x}) = -\sum_i h_i x_i - \sum_{i,j} x_i J_{ij} x_j \tag{3.1}$$

$$E_I(\mathbf{x}) = -\sum_i \bar{h}_i x_i \tag{3.2}$$

In other words, it is an energy that only considers bias parameters, denoted by $\bar{h}_i$. Note that while these parameters represent the same conceptual property (i.e. a neuron's tendency to fire), they will not be exactly equivalent in practice due to the fact that the independent model does not have the expressive power to account for coincidence rates between neurons and as a result will absorb these effects into the $\bar{h}_i$ parameters. However, the discrepancies between the two are often small and therefore will be assumed to be approximately equivalent in the following formulation ($h_i \approx \bar{h}_i$). The pairwise probability distribution can then be divided by the corresponding independent distribution, giving:

$$\frac{P_{pair}(\mathbf{x})}{P_I(\mathbf{x})} = \frac{Z_I}{Z_{pair}} \frac{e^{-E_{pair}(\mathbf{x})}}{e^{-E_I(\mathbf{x})}} \tag{3.3}$$

Where Z_{pair} and Z_I are the normalisation factors of the pairwise and independent distributions respectively. Considering only the exponential terms, under the previous approximation, it follows that the field parameters cancel out, leaving only the interaction terms:

$$\frac{e^{-E_{pair}(\mathbf{x})}}{e^{-E_I(\mathbf{x})}} = exp(\sum_i h_i x_i + \sum_{i,j} x_i J_{ij} x_j - \sum_i \bar{h}_i x_i) \tag{3.4}$$

$$\frac{P_{pair}(\mathbf{x})}{P_I(\mathbf{x})} = \frac{Z_I}{Z_{pair}} e^{\sum_{i,j} x_i J_{ij} x_j} \tag{3.5}$$

Renormalising this numerically gives $P_{H0}(\mathbf{x})$, a distribution with the effect of biases completely removed.

$$P_{H0}(\mathbf{x}) = \frac{1}{Z_{H0}}\left(\frac{P_{pair}(\mathbf{x})}{P_I(\mathbf{x})}\right) = \frac{e^{\sum_{i,j} x_i J_{ij} x_j}}{Z_{H0}} = \frac{e^{-E_{H0}}}{Z_{H0}} \tag{3.6}$$

Using this distribution it is then possible to scale the original pairwise probability and renormalise to obtain further distributions, as follows:

$$(P_{H0}(\mathbf{x}) * P_{pair}(\mathbf{x})) = \frac{e^{-E_{H0}}}{Z_{H0}} \frac{e^{-E_{pair}}}{Z_{pair}} \tag{3.7}$$

Considering again only the exponential terms, it is clear that the interaction terms can be grouped together:

$$e^{-E_{H0}(\mathbf{x})} e^{-E_{pair}(\mathbf{x})} = exp(\sum_{i,j} x_i J_{ij} x_j + \sum_i h_i x_i + \sum_{i,j} x_i J_{ij} x_j) \tag{3.8}$$

$$(P_{H0}(\mathbf{x}) * P_{pair}(\mathbf{x})) = \frac{exp(\sum_i h_i x_i + 2\sum_{i,j} x_i J_{ij} x_j)}{Z_{H0} Z_{pair}} \tag{3.9}$$

Again, renormalising this numerically gives a further distribution, $P_{2J}(\mathbf{x})$:

$$P_{2J}(\mathbf{x}) = \frac{1}{Z_{2J}} (P_{H0}(\mathbf{x}) * P_{pair}(\mathbf{x})) = \frac{exp(\sum_i h_i x_i + 2\sum_{i,j} x_i J_{ij} x_j)}{Z_{2J}} = \frac{e^{-E_{2J}}}{Z_{2J}} \tag{3.10}$$

This procedure can then be repeated, effectively introducing an additional parameter α amplifying the interaction terms in the energy function, leading to a set of distributions with energy functions given by:

$$P_{\alpha J}(\mathbf{x}) = \frac{1}{Z_{\alpha J}} (\alpha P_{H0}(\mathbf{x}) * P_{pair}(\mathbf{x})) = \frac{e^{-E_{\alpha J}}}{Z_{\alpha J}} \tag{3.11}$$

$$E_{\alpha J}(\mathbf{x}) = -\sum_i h_i x_i - \alpha \sum_{i,j} x_i J_{ij} x_j \tag{3.12}$$

Where the parameter α boosts the probabilities of patterns that arise as a result of the interaction terms, J. The aim of this is to remove patterns that can be explained by independent firing and give extra weight to multi-neuron patterns that are observed with low probabilities in a hope to emphasise changes in correlation structure within a network. Due to the renormalisation requiring the calculation of a partition function, only small networks (10 neurons) were considered for analysis. For models of this size

the renormalisation would give the effect of removing the silent or '-1' pattern but for larger networks, say 1000 neurons, this pattern may not be frequently observed and other patterns would be removed.

In the statistical mechanics formulation of this model there is an additional parameter β that is proportional to the inverse of the temperature of the physical system in question. The parameter scales the energy ($P = \exp(-\beta E)$) and for simplicity, the models used in statistical analysis have β set equal to 1 [8, 10]. However, adjusting this parameter has a significant effect on the energy landscape of the patterns. As β is increased (temperature goes down), the troughs and peaks of the energy landscape sharpen until they reach what is known as a 'frozen' state in the limit $\beta \to \infty$ (equivalent to the stationary distribution of a Hopfield network [85]). In this state the model is only able to inhabit a low number of observable states. In a similar manner, when scaling the interaction terms above the energy landscape will approach a state in which only a number of patterns are likely to be observed, but these patterns will arise as a result of high interaction terms rather than a firing rate effect.

3.2.2 Simulated Data - Dichotomised Gaussian Models

In order to explore the potential applications of the Ising model renormalisation, spiking data was simulated using Dichotomised Gaussian (DG) models. DG models offer a simple and effective method for simulating spiking networks by directly specifying the firing rate distribution and correlation structure [86]. Additionally, despite being purely synthetic, DG networks have been shown to successfully reproduce many characteristics of high-order correlations found in biological neuronal networks [87, 88, 89].

For a network of N neurons, the aim is to generate discrete spike trains of bin size t, $X_i(t) = 0$ or 1, by sampling from a joint distribution of N correlated Bernoulli random variables with means $\mu_i = \langle X_i(t) \rangle_t$ and covariances $\Sigma_{ij} = \left\langle X_i(t) X_j(t) \right\rangle_t - \langle X_i(t) \rangle_t \left\langle X_j(t) \right\rangle_t$. This is done by first sampling from a N-dimensional Gaussian random variable U and thresholding to 0 and 1:

$$X_i = 1 \text{ iff } U_i > 0 \text{ where } U \sim N(\gamma, \Lambda) \tag{3.13}$$

Due to the applied thresholding X will not have the same mean and covariance as U and so this discrepancy must be corrected for. The desired mean μ and covariance σ are related to γ and Λ through the following equations:

$$\mu_i = \Phi(\gamma_i) \tag{3.14}$$

$$\Sigma_{ii} = \Phi(\gamma_i)\Phi(-\gamma_i) \tag{3.15}$$

$$\Sigma_{ij} = \Psi(\gamma_i, \gamma_j, \Lambda_{ij}) \tag{3.16}$$

where $\Psi(x, y, \lambda) = \Phi_2(x, y, \lambda) - \Phi(x)\Phi(y)$, Φ is the cumulative distribution of a univariate Gaussian with mean 0 and variance 1, and $\Phi_2(x, y, \lambda)$ is the bivariate counterpart with correlation λ. Hence the desired firing parameters can be calculated numerically by inverting the above equations.

Using this model allows for a controlled simulation of networks with desired firing rates and correlations, enabling an ideal dataset for understanding the workings and benefits of the Ising model renormalisation.

3.3 Results - Simulated Data

3.3.1 Network Properties

Using the method detailed in Sect. 3.2.2 a baseline dataset was generated using sets of parameters for mean and covariance. For more biological realism, neuron rates were drawn from a lognormal distribution with a given mean and variance while their covariances were set to be normally distributed with values close to 0. This was in accordance with evidence that neuronal firing rates exhibited such distributions in numerous areas across the brain [90, 91, 92, 93]. These distributions were parameterised by $\bar{\mu}^{(i)}$ and $\bar{\Sigma}^{(i)}$, which denote the means of the rate and covariance distributions respectively.

Starting from the baseline values $(\bar{\mu}^{(1)}, \bar{\Sigma}^{(1)})$, the means of the two distributions were then systematically increased in turn (while keeping the other constant) in order to generate an array of subsequent populations to be compared to the original (Fig. 3.1).

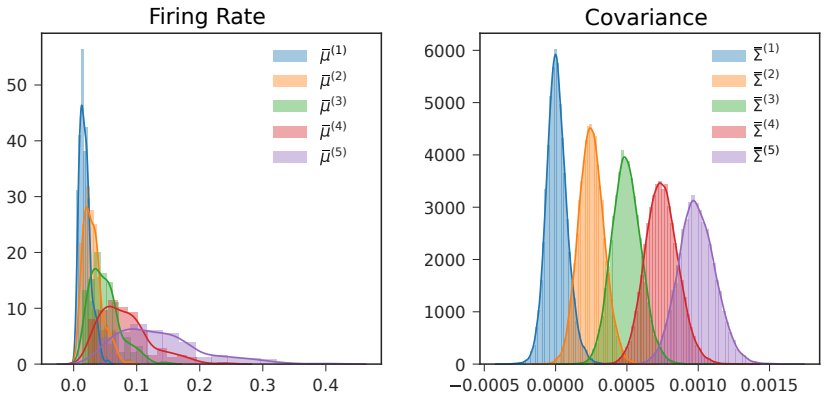


Figure 3.1: (a) Firing rate and (b) covariance distributions of different simulated datasets (corresponding parameters in legend).

For completeness, a dataset was generated at each combination of parameter values, hence creating an 5×5 array of datasets. This can be further visualised in Fig. 3.2, where the mean firing rate, covariance and correlation is computed from the generated spike trains and plotted per dataset. Note here that the correlation (ρ_{ij}) between two neurons is related to the covariance through $\rho_{ij} = Cov(i,j)/\sqrt{Var(i)Var(j)}$. For Bernoulli random variables, variance is equal to $p(1-p)$, where p is the probability of being active at a given time (i.e. the firing rate). This direct link between correlation, rate and covariance can be observed in the generated datasets.

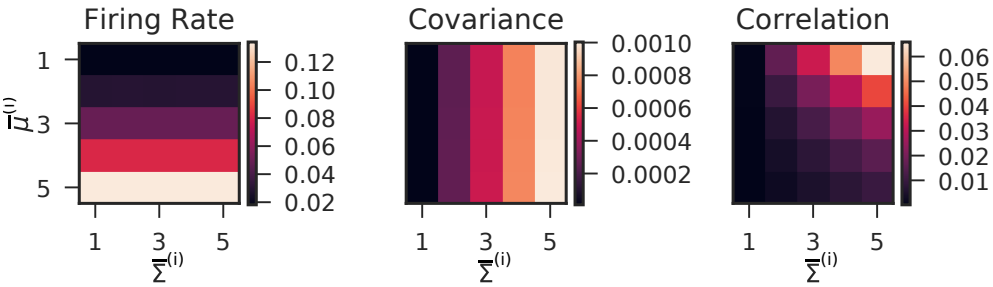


Figure 3.2: Demonstration of dataset generation. Each entry in the surface plots corresponds to a dataset parameterised by μ and Σ. From left to right entries show the mean firing rate, covariance and correlation of the simulated networks.

Network activity generated by a number of different DG models can be seen in Fig. 3.3-3.4, where the mean rate and mean covariance are increased respectively. The

effect of increasing these network parameters is clearly demonstrated in the observed frequency and synchrony of the spiking activity.

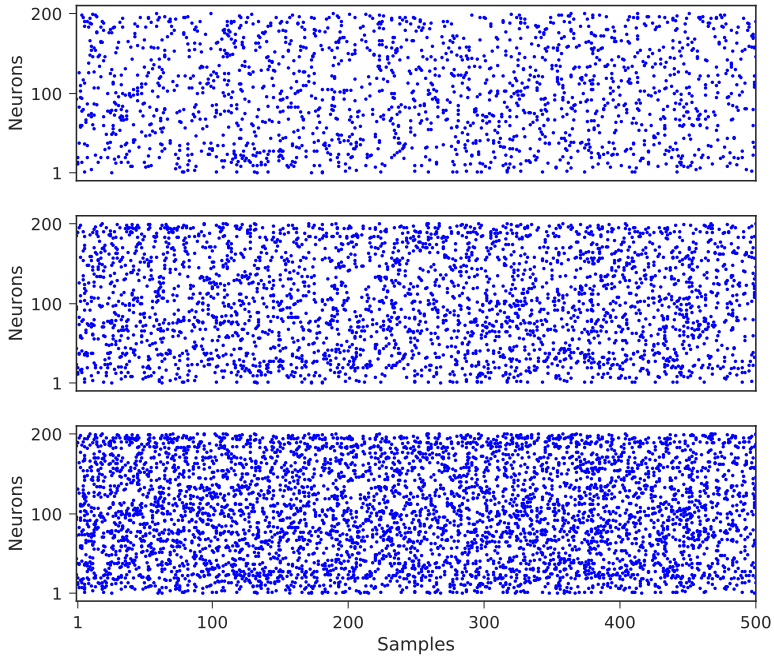


Figure 3.3: Raster plots of simulated spike train from Dichotomised Gaussian models as the mean rate is increased (top to bottom)

By generating networks with a systematic range of governing parameters, the idea is to create a set of simulated datasets that can be compared to real life experiments in which the aim is to detect changes in network structure over time. For example, to examine a network remodelling in reaction to external chemicals such as in the case of neuronal cultures, or in response to an experimentally induced plasticity regime in awake behaving animals. As the rates and correlations in this case are controlled, it offers an ideal environment for testing the rescaled model's effectiveness in detecting changes in network correlation structure.

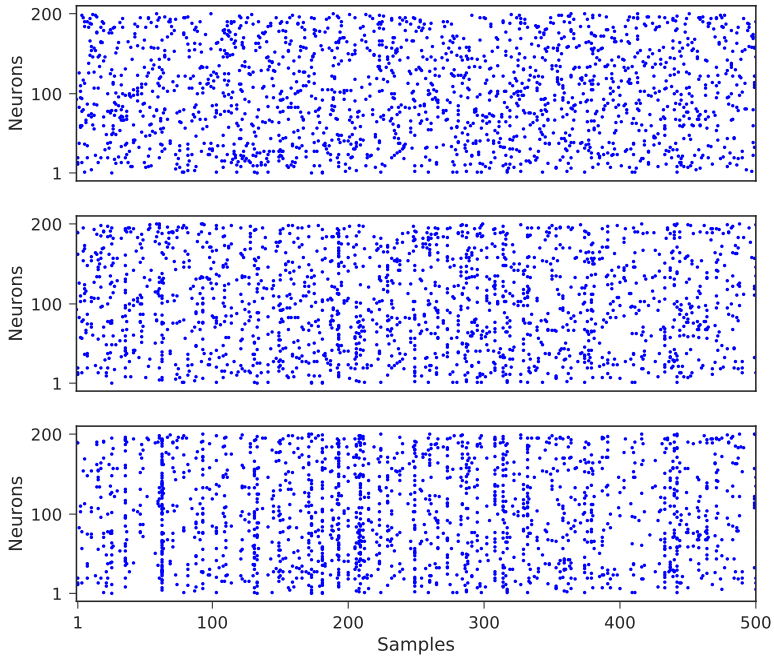


Figure 3.4: Raster plots of simulated spike train from Dichotomised Gaussian models as the mean covariance is increased (top to bottom)

3.3.2 Model Fits and Likelihood

A first test of model fit is to plot the pattern probabilities predicted by the model against those calculated from the DG model. The DG pattern probabilities were calculated using the cumulative density function of the generating multivariate normal distribution, which was then converted into a probability mass function using the inclusion-exclusion principle [94, 95]. As can be seen in Fig. 3.5, the models predict patterns well but uncertainty grows for estimates of patterns that occur rarely within the dataset (i.e. low probability patterns). Furthermore, as the covariance of the data is increased, the improvement of the Ising model over the independent becomes clearer.

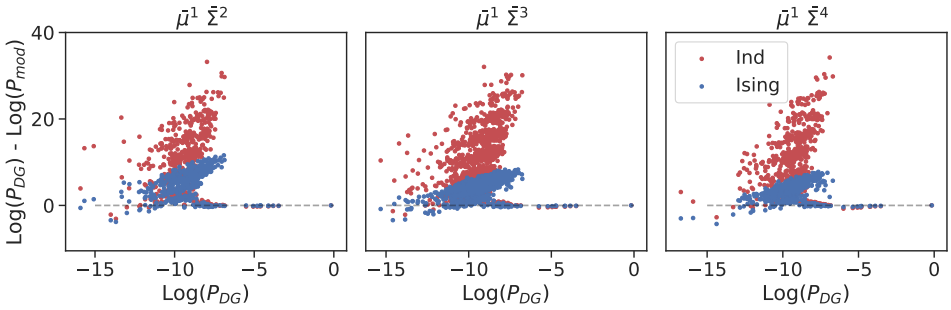
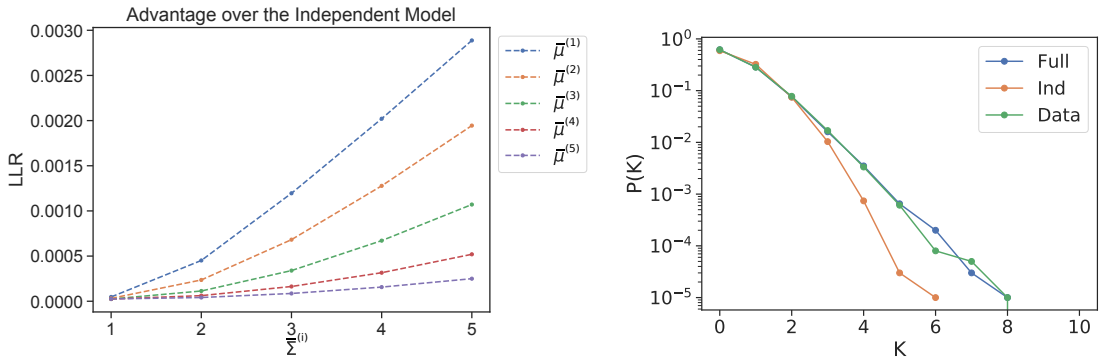


Figure 3.5: Log probabilities predicted by the Ising and independent models vs. those observed calculated from the DG model, for three different datasets of increasing covariance.

Fig.3.6(a) shows the log-likelihood ratio between the Ising model and the independent model for increasing values of Σ at each starting μ. This ratio compares the log-likelihoods of the pairwise and independent models under the observed data, and is given by the formula:

$$LLR = \langle \log(P_{Model}) - \log(P_{Ind}) \rangle_{Data} \tag{3.17}$$

This is effectively a measure of how much better the Ising model explains the data than the baseline case of the independent model. The plot shows a general increase in log likelihood ratio as covariance increases, an intuitive result as for low values of Σ there is less synchronous activity and a lower amount of higher order correlations for the model to capture (independent model is sufficient to explain the data). This advantage becomes gradually more significant as covariance increases and the Ising model has substantially more higher order statistics to capture.



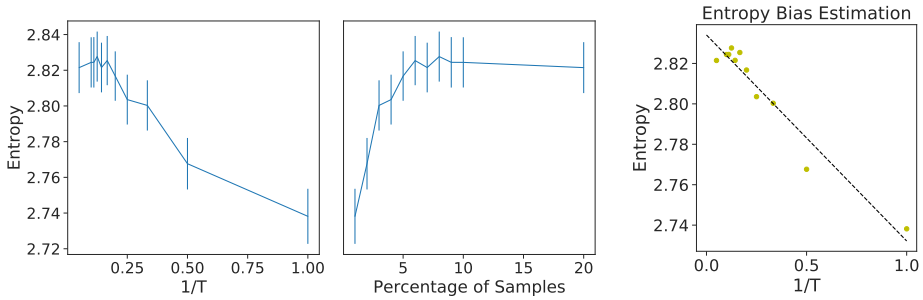
(a) (b)

Figure 3.6: (a) Log likelihood ratio for each simulated dataset. (b) Predicted synchrony distribution P(K) for the independent and full model, shown alongside the observed P(K) from an example dataset

Additionally, as μ in increased, the effect of increased covariance is weaker. This is due to the fact that as rates become higher they begin to dominate the effect of covariance on the pattern distribution, leading to only small increments in LLR with higher covariance. A further example of the Ising model's superior fit can be seen in Fig. 3.6(b), where the predicted synchrony distribution is shown for both models against the observed distribution from an example dataset.

3.3.3 Entropy Estimation

When fitting Ising models it is important that an adequate amount of data is used to obtain a good fit. This is to ensure that networks are observed over enough time to ensure a sufficient proportion of their underlying pattern distribution is realised. When models are fit on insufficient data there is an inherent bias in the entropy estimated by the model. In order to test whether a sufficient amount of data has been acquired, models were fit on subsets of the data for increasing values of T and entropy was calculated at each. Fig. 3.7 shows the relation between entropy and T, demonstrating that the entropy levels off at around $\sim 10\%$ of the original $T = 100000$ samples, indicating that the dataset is sufficiently large. However, it is important to note that the entropy also depends on other factors such as the network size N and so the 10% value is not a hard threshold for sufficient data.



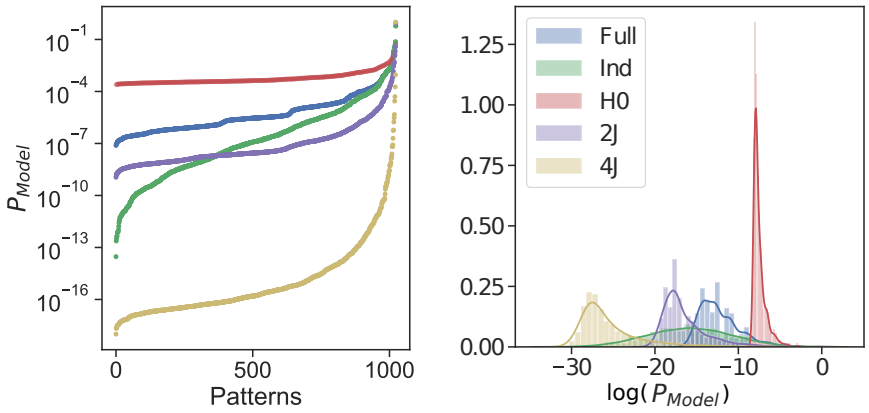
(a) (b)

Figure 3.7: Calculated model entropy as a function of dataset length T and $1/T$. Also shown, a linear fit of the relation between entropy and $1/T$, allowing for an extrapolation of $T \to \infty$

To obtain an unbiased entropy measure, a linear model was fit to the relation between entropy and $1/T$, making it possible to extrapolate to $T \to \infty$ (the origin). Entropy at this value offers an estimate of the entropy calculated on a dataset with infinite samples.

3.3.4 Exploration of Renormalised Distributions

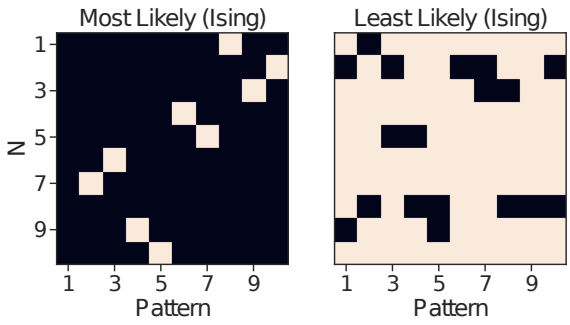
Before proceeding with further analysis using the rescaled Ising models it is important to first observe the effects of this modification on the model's pattern distribution. In removing the firing rate contribution completely (P_{H0}), each pattern now arises entirely from interaction terms, with neurons that have stronger J_{ij} terms determining the most likely patterns. In order to explore these effects, the following section details a number of plots that all correspond to models fit to a particular simulated dataset ($\bar{\mu}^{(3)}, \bar{\Sigma}^{(5)}$). By plotting P_{H0} alongside P_{Ising} (both ranked) it is evident that the distribution as a whole is more uniform, and that the majority of patterns become more likely as a result, Fig. 3.8.



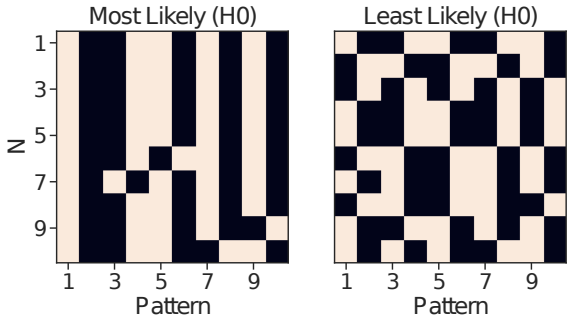
(a) (b)

Figure 3.8: (a) Ranked log probabilities of different models. (b) Corresponding distribution plots

A further test to investigate how these distributions differ is to see which patterns the models predict to occur the most/least frequently. Fig. 3.9 demonstrates this difference clearly: for the Ising model, the most likely patterns are the silent pattern (all -1's) and the single spike patterns, arising as a result of the bias parameters, whereas the least likely patterns are those with a high number of co-active neurons ($k = 8, 9$). However, in the H0 model the most likely patterns are those with the highest number of co-active or co-inactive neurons: all -1's or all +1's, followed by those with $(N-1)$ -1's or +1's and so on, whereas the least likely patterns are those with half -1 and half +1. Note that while some of the most likely patterns are the same in the two models, they are occurring as a result of different effects.



(a)



(b)

Figure 3.9: Ten patterns with the highest (left) and lowest (right) predicted probabilities for the (a) Ising and (b) H0 model

Further to this result, plotting the pattern probabilities as a function of the number of active units k also reveals a symmetry in the distribution, Fig. 3.10. This symmetry is a result of the $[-1, 1]$ representation of the model - without any bias, each pattern and its mirror (pattern with flipped components) will have identical probabilities.

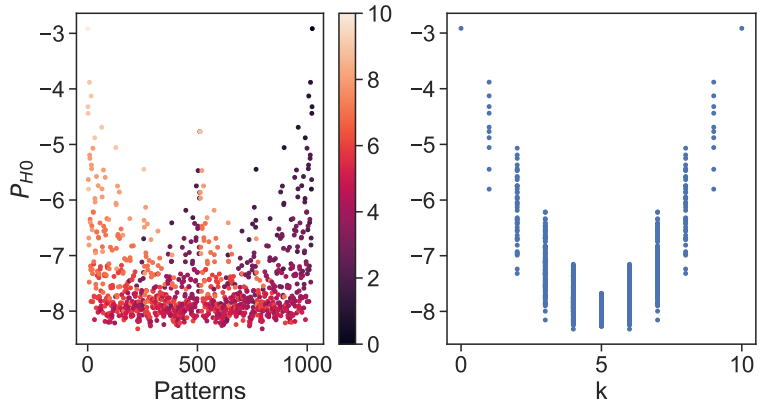


Figure 3.10: H0 model log probability plot coloured by k values (left) and ranked by k values (right)

Looking at the distribution over half of the patterns there is a clear relation between the number of active units in a pattern and that pattern's probability. This follows from the above reasoning that patterns with a higher number of co-active (or co-inactive) units will have a higher probability. A scatter plot of P_{Ising} versus P_{H0} also reveals effective shift of each pattern probability under model renormalisation.

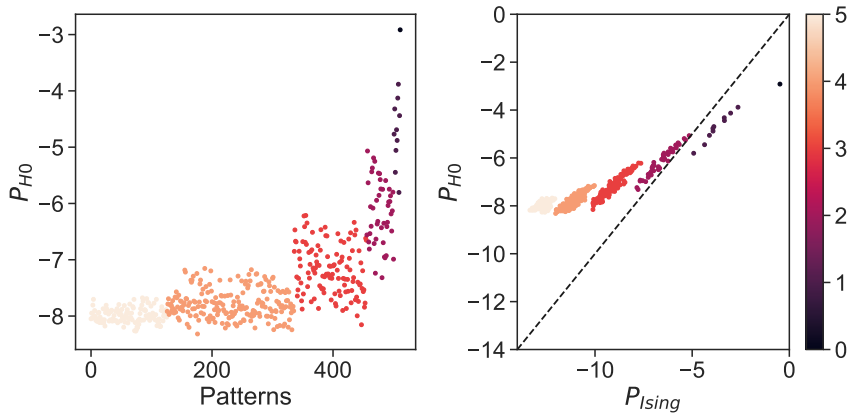


Figure 3.11: One half of the symmetric H0 model log probabilities, coloured by k (left). These same probabilities scattered against the corresponding pattern probabilities predicted by the Ising model (right)

A further exploration of $P_{H0}(\mathbf{x})$ can be carried out by sampling a sufficiently large spike train from the distribution, as is shown in Fig. 3.12. Interestingly with the effect of firing rate completely removed, the computed mean rates of each synthetic spike train fall around the same value, $\langle x_i \rangle \approx 0.5$. This is effectively assigning a probability of $p(x_i = 1) \approx 0.5$ for each neuron to be active in a given time sample where the discrepancies around 0.5 are determined by the correlation structure dictated by the original model, creating a sort of surrogate distribution. In the absence of any correlation structure, this would result in a completely uniform distribution across patterns.

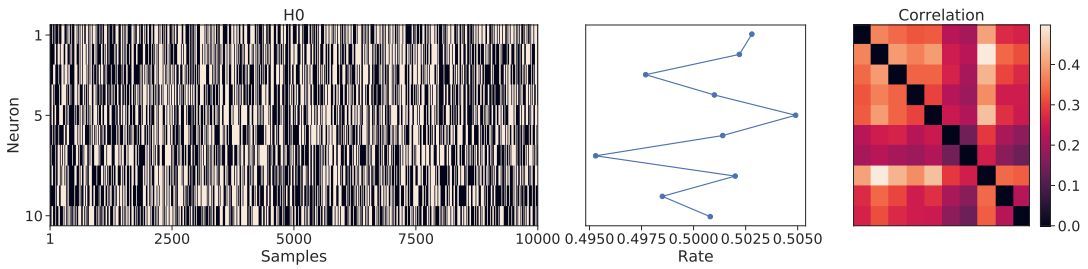


Figure 3.12: Example spike train sampled from P_{h0} (left). Neuron rates and correlations computed from this sample spike train (middle & right)

In the intermediate case, when the model is rescaled to boost the effects of interactions but keeps its bias terms $(P_{\alpha J})$, patterns that arise mainly due to the J parameters will become more likely, Fig. 3.8. If α continues to increase then an effect similar to that of increasing β in the original Ising model is observed. What happens here is that the temperature leads to a sharpening of the model energy landscape, eventually leading to a 'frozen' state in which only a small number of patterns can be observed [20]. In contrast here, when scaling the J parameters, the energy landscape is altered in such a way that only the patterns that are largely influenced by interaction terms are affected.

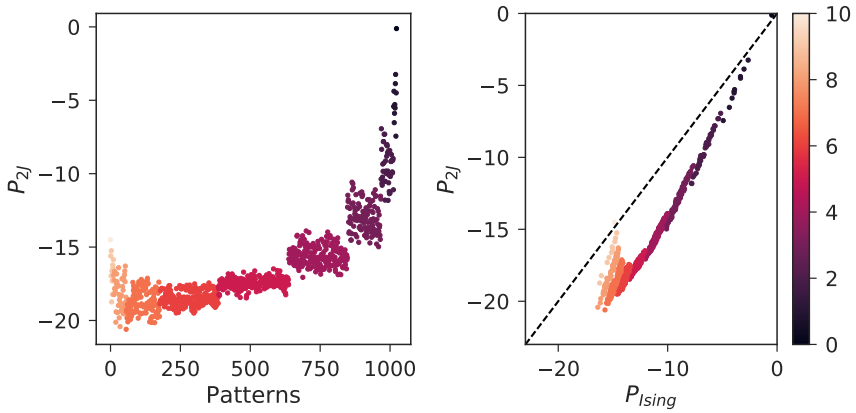


Figure 3.13: 2J model log probabilities, coloured by k (left). These same probabilities scattered against the corresponding pattern probabilities predicted by the Ising model (right)

Fig. 3.13 shows the distributions P_{2J}, again ranked by k. Here it is evident from the shift in the tail of the distribution that patterns with many coactive neurons become more likely, and that the ordering of least likely patterns begins to change based on the interaction parameters, Fig. 3.14.

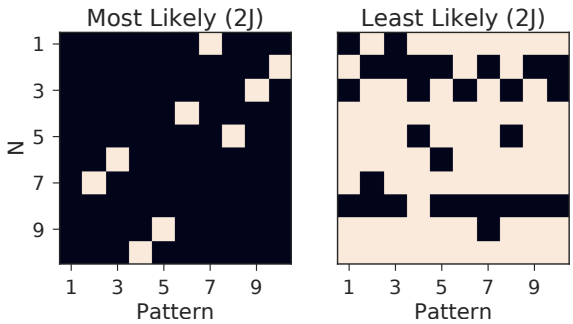


Figure 3.14: Ten patterns with the highest (left) and lowest (right) predicted probabilities for the 2J model

Sampling from this model maintains (roughly) the same firing rate distribution as the original model, with a slight boost in correlation structure, Fig. 3.15.

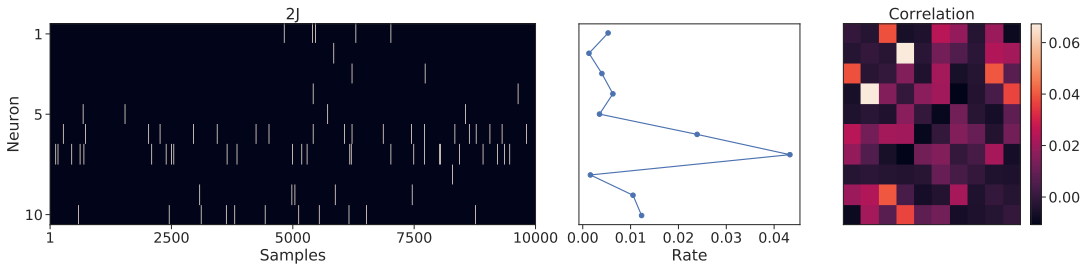


Figure 3.15: Example spike train sampled from P_{2J} (left). Neuron rates and correlations computed from this sample spike train (middle & right)

3.3.5 Correlation Changes in DG Models

To utilise the renormalised Ising distributions the next step is to identify the types of analyses to be carried out on these distributions in order for them to be informative. The most obvious choice is a divergence measure between the distributions at different points in 'time' (i.e. as the DG parameters underlying their simulation are modified).

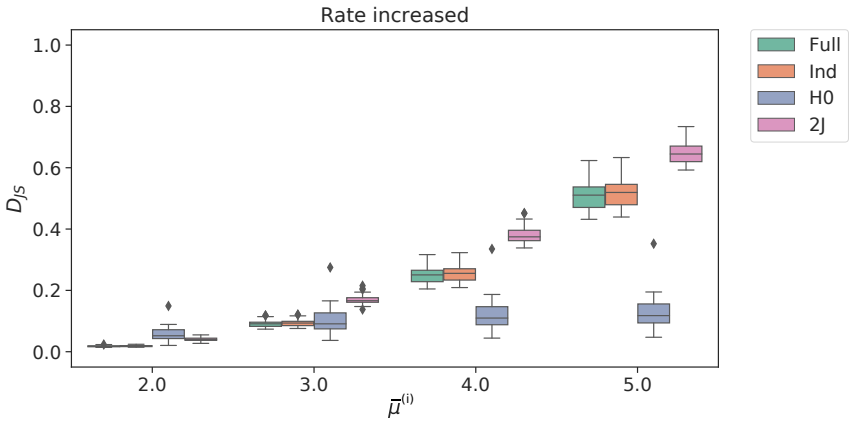


Figure 3.16: Jenson-Shannon divergence between the baseline ($\bar{\mu}^{(1)}$,$\bar{\Sigma}^{(1)}$) and subsequent model distributions as $\bar{\mu}^{(i)}$ is increased across datasets

Starting with the case in which $\bar{\mu}^i$ is altered across datasets, Jenson-Shannon divergence was computed between each dataset and the original. Results indicate that while divergence increases with $\bar{\mu}^i$ for the Ising and Independent model, it levels off to a constant value for the H0 model.

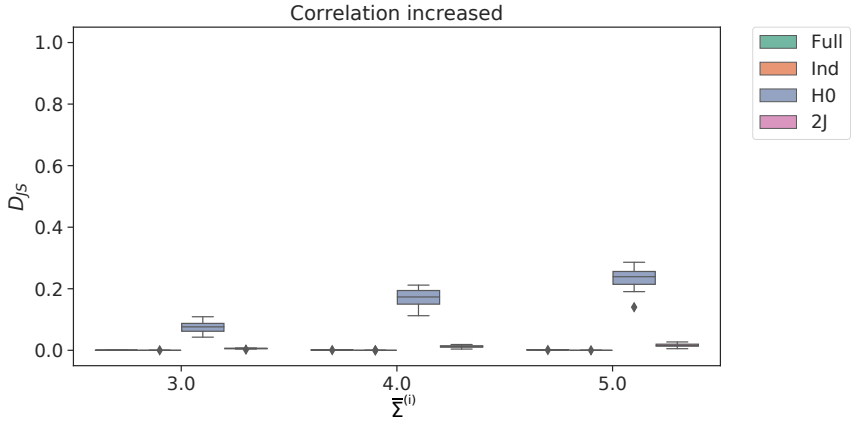


Figure 3.17: Jenson-Shannon divergence between the baseline ($\bar{\mu}^{(1)}$,$\bar{\Sigma}^{(1)}$) and subsequent model distributions as $\bar{\Sigma}^{(i)}$ is increased across datasets

Alternatively, when only $\bar{\Sigma}^{i}$ is increased, the H0 model is the only model to exhibit a significant difference between datasets, hence highlighting a situation in which this model can be informative beyond the original one. When both $\bar{\mu}^{i}$ and $\bar{\Sigma}^{i}$ are increased, all models give a divergence across cases. However due to the slight increase of the H0 divergence it is indistinguishable from the case for which only $\bar{\mu}^{(1)}$ is altered.

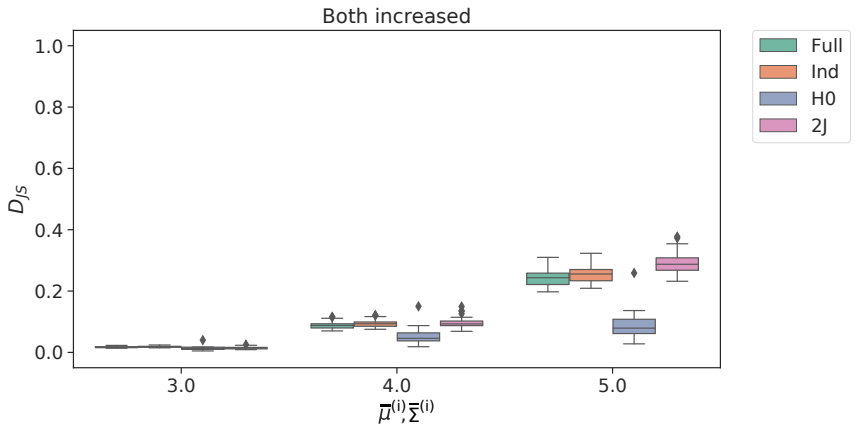


Figure 3.18: Jenson-Shannon divergence between the baseline ($\bar{\mu}^{(1)}$,$\bar{\Sigma}^{(1)}$) and subsequent model distributions as both $\bar{\mu}^{(i)}$ and $\bar{\Sigma}^{(i)}$ are increased simultaneously across datasets

In real networks, a case in which correlation changes while rates are held constant is unlikely to occur and therefore further analysis is necessary to isolate the effects of the H0 model. Such analysis can be carried out with the use of a surrogate dataset - a dataset that conserves the firing rates of the original but destroys all correlations through random shuffling of the spike trains across time. This type of analysis has

been vital in providing evidence for hypotheses relating observed distributions to optimal coding regimes [16, 50, 51]. The idea is to fit models to the original and surrogate distributions and compute a divergence measure between the two, in order to quantify the amount that correlation structure contributes to the overall firing pattern distribution. Fig. 3.19 demonstrates this effect clearly for each starting value of $\bar{\mu}^{(1)}$ and $\bar{\Sigma}^{(i)}$.

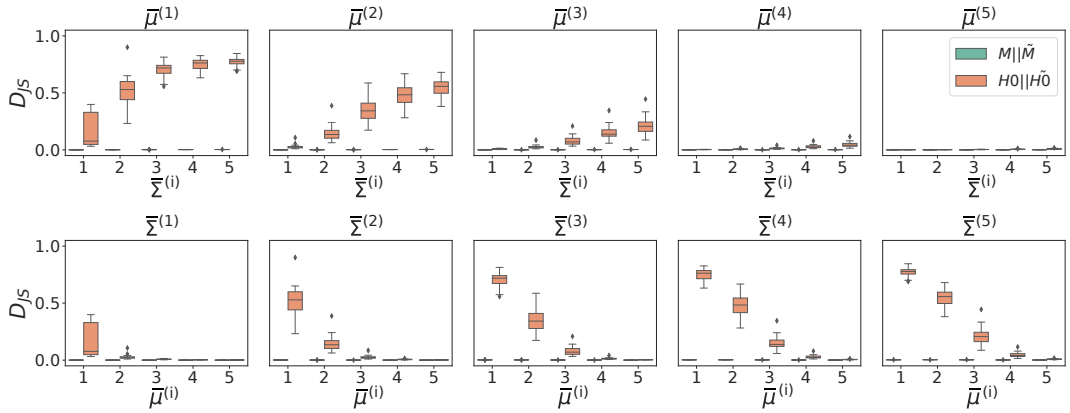


Figure 3.19: Jenson-Shannon divergence between models of original data and models of surrogate (shuffled) data

Results indicate the ability of the H0 model to capture underlying correlation structure that is otherwise masked by the effects of firing rate. As the $\bar{\Sigma}^{(i)}$ parameter is increased, the divergence between the original and surrogate H0 models grows large while the corresponding divergences of the Ising model remain very low. This can be said in most cases apart from the high rate regime, where the dominant effect of firing rate causes the result to break down, Fig. 3.20.

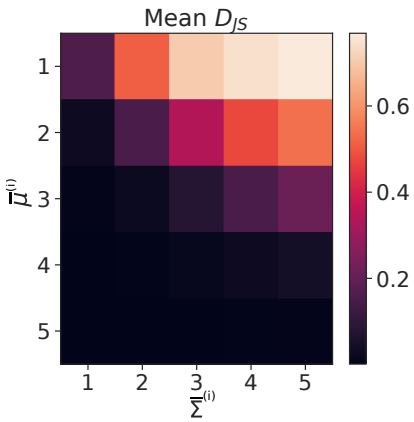


Figure 3.20: Mean Jenson-Shannon divergence of all 20 groups for all values of μ and Σ

To conclude, the H0 model offers an interesting insight into the correlation structure of simulated DG networks and enables further insight into how these structures change beyond the reach of the pure Ising model, where these subtle effects are often masked by the dominating changes of neural firing rates.

3.4 Results - MEA Recordings from Cultures

3.4.1 Population Activity

Mean spatial activity across experimental phases of chip 140 can be seen in Fig. 3.21 (see Appendix A.1 for corresponding plots of remaining experiments). Areas of high activity can be seen in localised hubs across the chip, with correlation also higher among these smaller subnetworks. As discussed in chapter 2, data extracted from these chips can be readily converted into binary matrices through spike detection algorithms [73] and subsequent binning procedures. The choice of binsize $\delta\tau$ should be such that it is big enough to capture synaptic interactions but small enough to avoid too many polysynaptic events. A common method for estimating this value is to compute a cross-correlogram between neuron pairs to approximate the characteristic time scale of interactions between neurons. Calculations estimated a bin size of $\delta\tau = 20ms$, in accordance with a similar study carried out on cultured hippocampal neurons using the same MEAs [39].

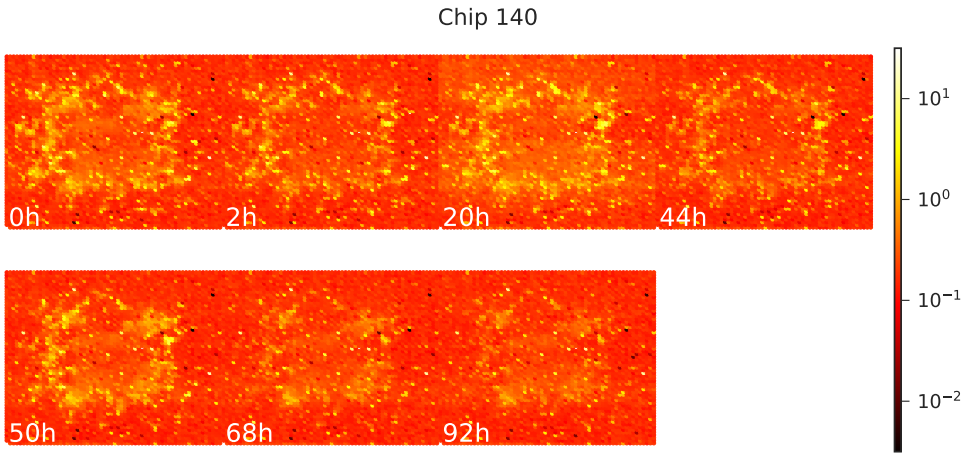


Figure 3.21: Spatial activity at different time points throughout the experiment

From the above recordings, groups of neurons were selected for fitting as follows: first of all neurons were thresholded by rate to remove neurons with extremely high or

low firing (between $0.1 - 10Hz$), a neuron was then selected at random and the closest 9 were taken to make up the full group. This was to ensure high correlation between selected neurons. Twenty groups of 10 neurons were subsequently sampled from the population using this procedure to ensure a sufficient fraction of the population was included for analysis. Each group was monitored and a modelled at each recording phase throughout the experiments.

3.4.2 Model Fits and Entropy Estimation

Again, model fits were investigated in order to validate further analysis. Predicted pattern probabilities were computed and compared against those observed from the datasets, as is shown for a single neuron group in Fig. 3.22.

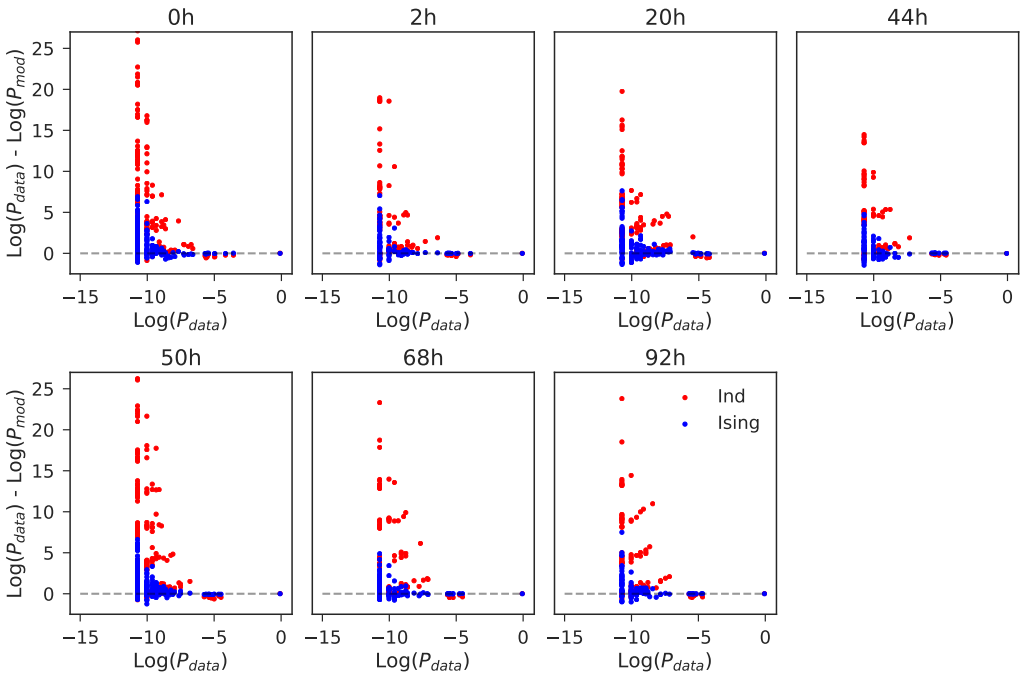


Figure 3.22: Predicted versus observed pattern probabilities for Ising and independent models, for a single neuron group within the chip 140 experiment

Pattern probabilities predicted using the independent model were also included to demonstrate the Ising model's ability to out-perform the baseline. This advantage is again demonstrated using the LLR, which is computed for each neuron group and averaged, Fig. 3.23.

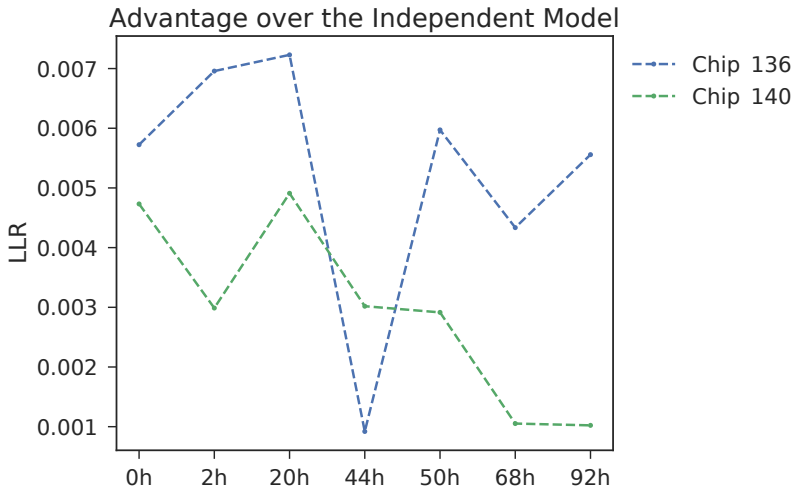


Figure 3.23: Average LLR over all groups across phases for chip 136 and 140

In order to quantify the model fit over all groups the Jensen-Shannon divergence between the predicted and observed pattern distributions was computed for each. Sufficiently low values were obtained indicating models had captured the observed distributions to within a tolerable level of indistinguishability, Fig. 3.24. If for some reason an adequate fit was not obtained for a certain group within a culture, these groups were discarded and a replacement group was sampled from the population.

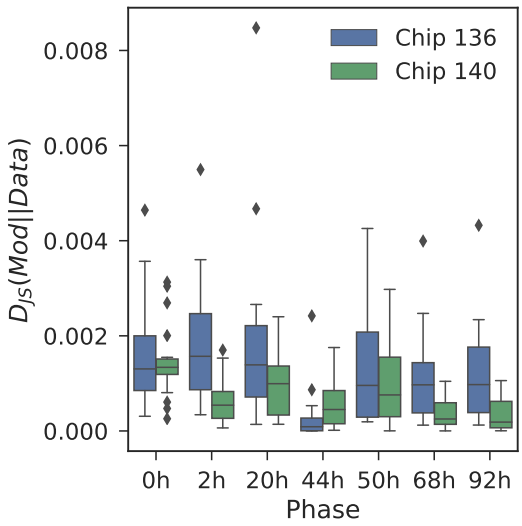
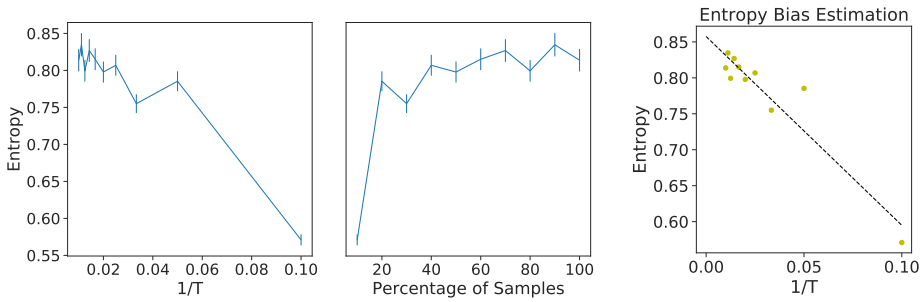


Figure 3.24: Jensen-Shannon divergence between model and observed pattern probabilities, for each group across all phases for chip 136 and 140

Additionally, models were fit to datasets of increasing length to ensure an acceptable amount of data was being considered. This was to verify whether the network

activity had enough time to express its pattern distribution to a sufficient degree. Fig. 3.25 shows the computed entropy of model pattern distributions as the time was increased, clearly levelling off at a value around 50% of the full recording length. Similar plots were found for all cultures considered for analysis therefore ensuring that a sufficient amount of data (full length recordings) were being considered.

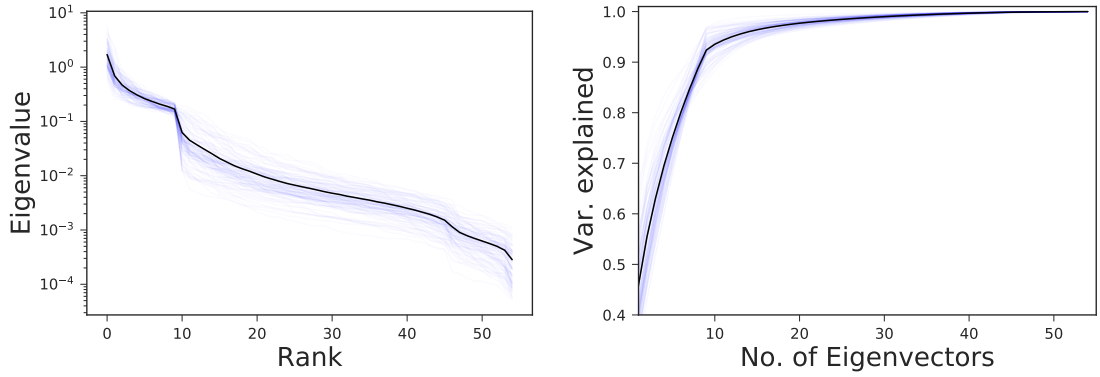


(a) (b)

Figure 3.25: Computed entropy for models fit to spiking activity of increasing length for chip 136

3.4.3 Eigenparameter Analysis

Having obtained satisfactory model fits, it is of interest to explore the eigendecomposition of the model FIM. As detailed in chapter 2, such analysis can identify important (sensitive) and unimportant (insensitive) dimensions in parameter space. These dimensions are named according to the magnitude of their influence on model behaviour. Fig. 3.26 shows the ranked eigenvalues of the FIMs fitted to all groups and all phases for chips 136 and 140. The span of the eigenvalues exhibits characteristics of a sloppy model due to its rapid decrease over an enormous range, therefore strengthening claims that sloppiness exists in spontaneously active neuronal networks [41].



(a) (b)

Figure 3.26: (a) Ranked eigenvalues of FIMs computed from Ising model fits to all groups and phases of chips 136 and 140 (shaded blue). Average shown in black. (b) Cumulative sum of the same (normalised) eigenvalues. Represents the amount of FIM variance explained as the number of subsequent eigenvectors is included

Such a characterisation also suggests that there exists certain 'stiff' (highly sensitive) and 'sloppy' (highly insensitive) dimensions in parameter space, with a small amount of stiff directions that largely dictate model behaviour and many sloppy ones that can undergo large change without significantly affecting the model distribution. This can be seen in the cumulative sum of normalised eigenvalues, in which the principal eigenvector explains on average 50% of FIM variance, hence specifying the stiffest direction in parameter space, whereas a high number of sloppy directions contribute $< 1\%$.

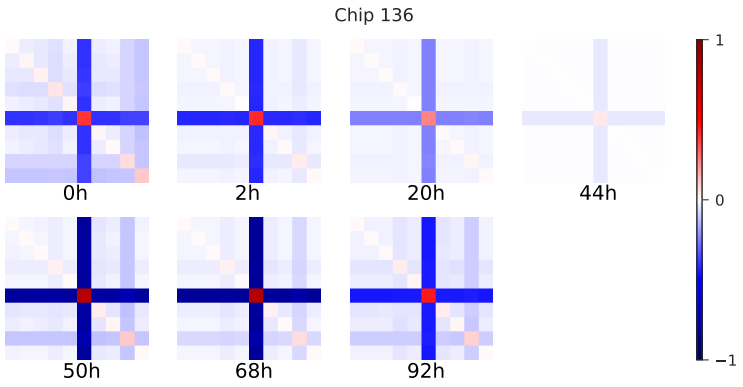


Figure 3.27: Principal eigenvectors scaled by the their eigenvalue for models fit to culture 136 across all experimental phases

Furthermore, looking at the principal eigenvectors for some of these sloppy models reveals a highly sparse structure and a strong stability across recording phases, as demonstrated in Fig. 3.27. The sparsity of the eigenvectors indicates that only a small number of parameter contribute to these stiff directions and more importantly these parameters belong to a small subset of neurons. Note here that while parameters will differ due to a change in firing rate and correlation over phases, the principal eigenvectors remain almost identical.

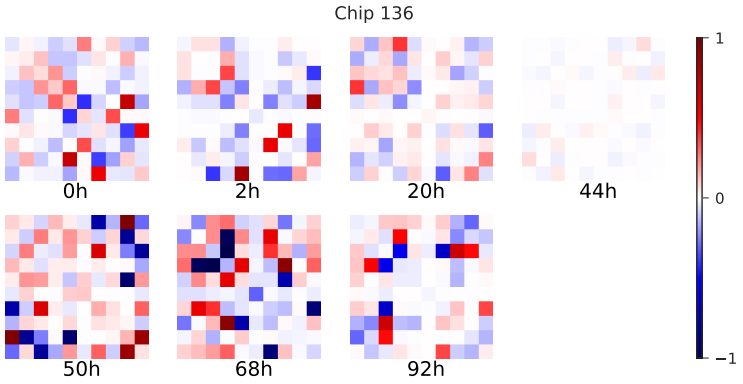


Figure 3.28: Eigenvectors corresponding to sloppy directions scaled by the their eigenvalue for models fit to culture 140 across all experimental phases. Note here the 30th eigenvector was selected.

Conversely, sloppy dimensions can be observed to exhibit a much denser structure, reflecting the higher number of parameters contributing to these dimensions, Fig. 3.28. Interestingly these eigenvectors are far less consistent across recordings reinforcing the idea that sloppy dimensions undergo large change without significantly affecting model behaviour.

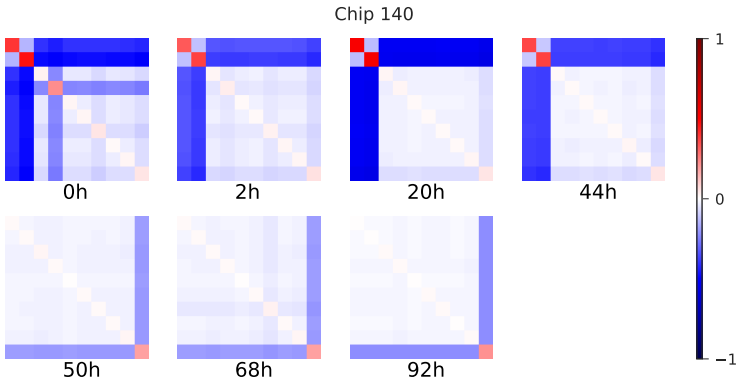


Figure 3.29: Principal eigenvectors scaled by the their eigenvalue for models fit to culture 140 across all experimental phases

These results are in line with findings that show stiff directions remain stable over time while sloppy dimensions are free to fluctuate, with these characteristics corresponding to a particular type of network mechanism in which a small subnetwork of neurons supports stability while allowing the full network to undergo remodelling without damaging network functionality [41]. Interestingly, while this same sloppy behaviour is observed for chip 140, the principal eigenvectors were found to be less consistent after washout, Fig. 3.29, indicating a permanent remodelling effect as a result of CNQX addition, which will be investigated further in the following section.

3.4.4 Neuronal Cultures Undergo Permanent Remodelling Effect

Mean firing rate and correlation for each selected group within the culture recordings 136, 140, 77 and 184 are shown in Fig. 3.30-3.32.

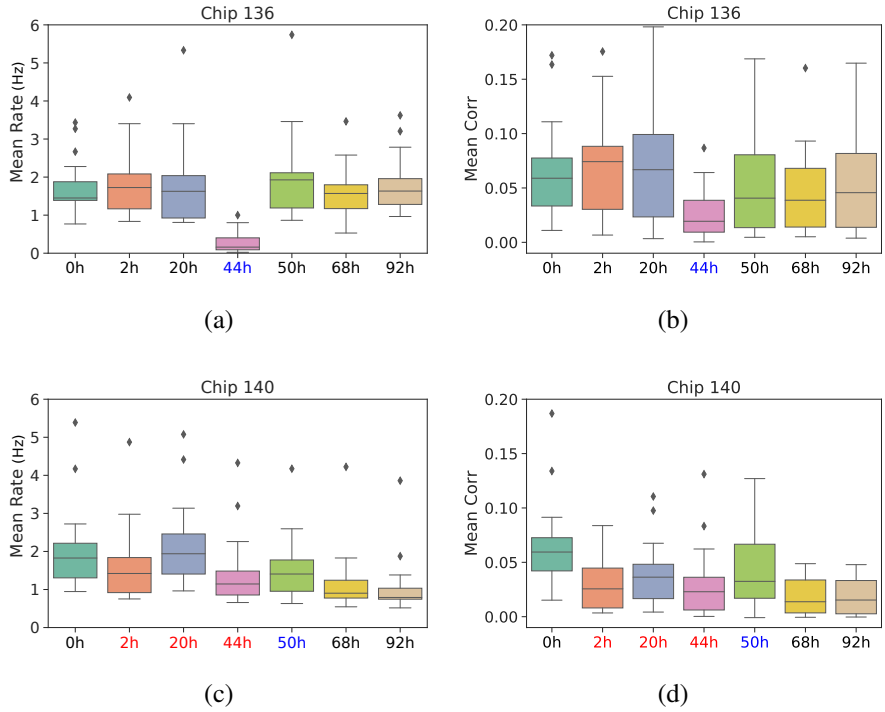
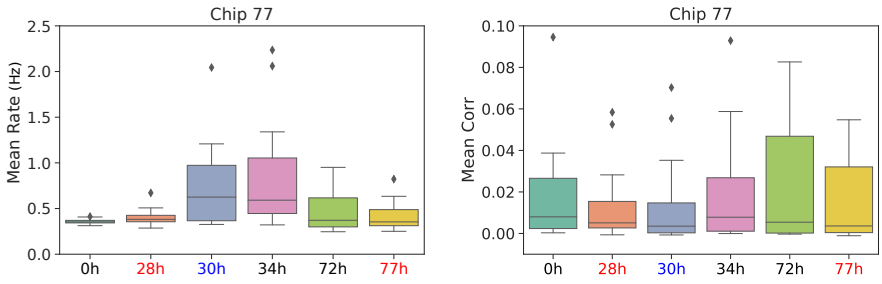


Figure 3.30: Mean firing rate and pairwise correlation for all selected groups within cultures 136 and 140, x-label colours reflect the type of recording phase, corresponding to the colours in Table. 2.1

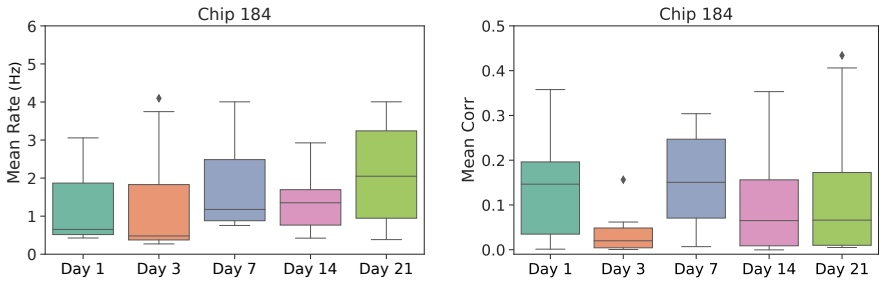
Overall, cultures maintain a relatively stable level of activity across recording with some fluctuations around baseline. The only outlier is in the case of culture 136 at the 44 hour mark, where the observed low activity is due to a medium change.



(a) (b)

Figure 3.31: Mean firing rate and pairwise correlation for all selected groups within culture 77, x-label colours reflect the type of recording phase, corresponding to the colours in Table. 2.1

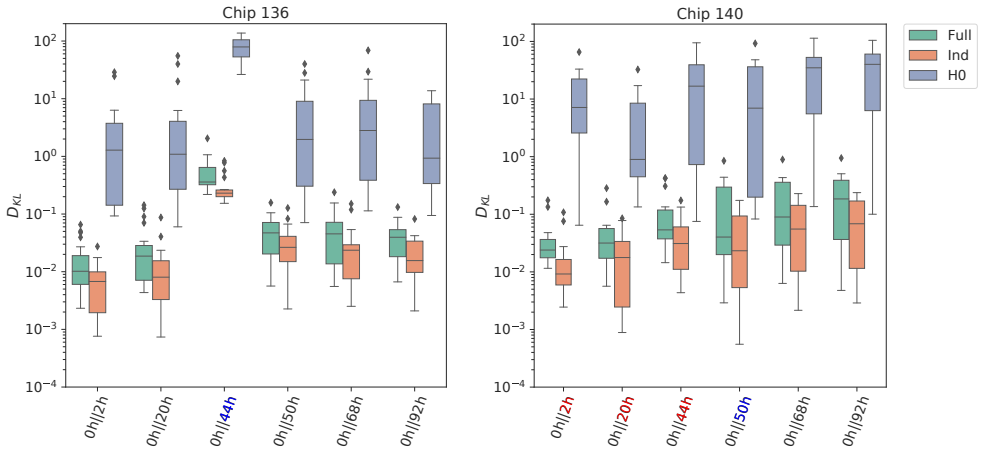
As expected, due to the inhibitory effects of CNQX, cultures that underwent chemical treatment experienced a larger drop in activity across the experiment period compared to those that did not.



(a) (b)

Figure 3.32: Mean firing rate and pairwise correlation for all selected groups within culture 184

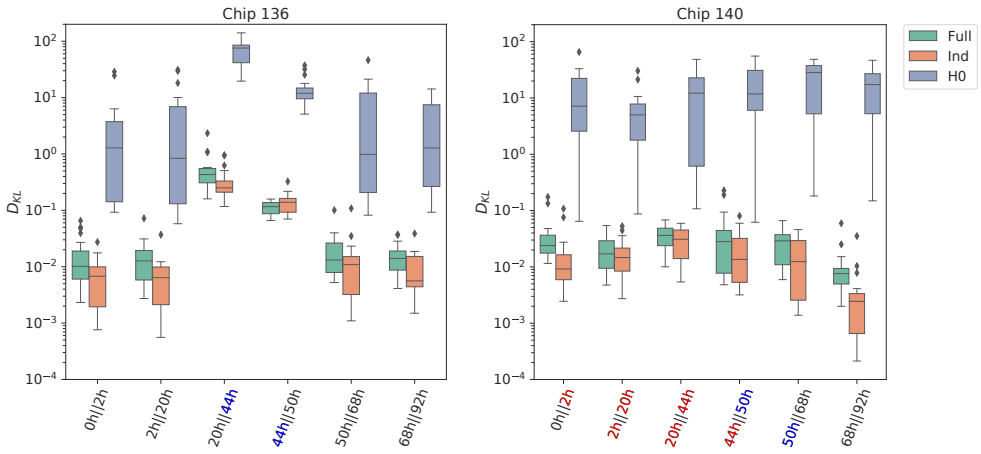
Cultures 136 and 140 allow for a simple and direct comparison due to the similarity of their age and recording protocol, with culture 136 acting as a control experiment in the absence of CNQX. As expected, the activity of culture 140 gives a larger divergence from the baseline recording compared to culture 136, as demonstrated in the KL divergence computed between three different distributions, Fig. 3.33.



(a) (b)

Figure 3.33: KL divergence between first and subsequent recording phases for cultures 136 and 140, using Ising (full), independent and H0 models. X-label colours reflect the type of recording phase, corresponding to the colours in Table. 2.1

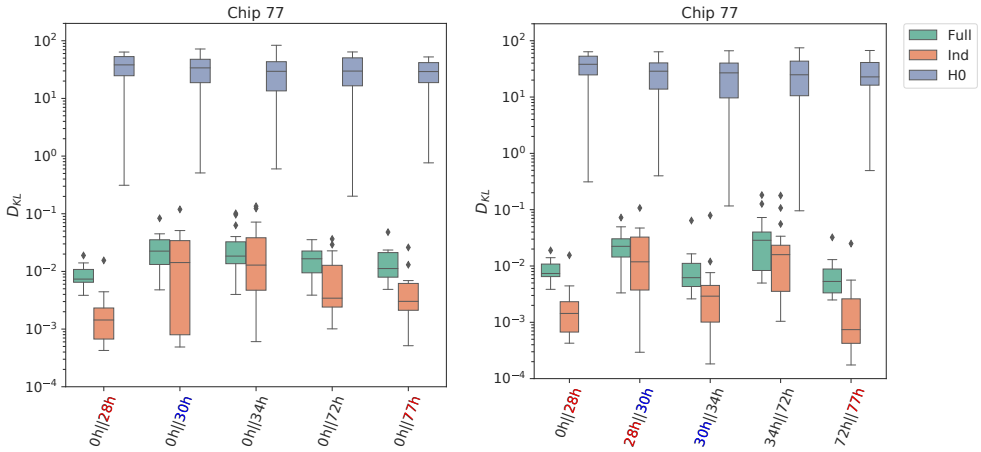
Interestingly, this divergence remains higher even after the washout, indicating a permanent remodelling effect that is not reversible, supporting findings from the eigenparameter analysis. The lower divergences for culture 136 also support the idea that stiff directions are stable across these recording phases, in the absence of induced inhibition.



(a) (b)

Figure 3.34: KL divergence between successive recording phases for cultures 136 and 140, using Ising (full), independent and H0 models. X-label colours reflect the type of recording phase, corresponding to the colours in Table. 2.1

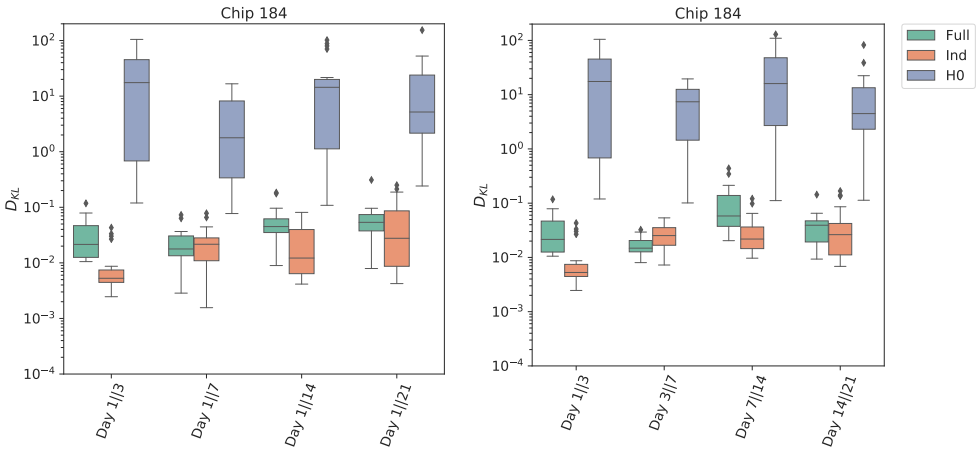
Culture 77 had a different recording protocol involving two separate CNQX injections, with the second dose being stronger than the first. In agreeance with previous results, all three model divergences are higher across phases as a result of induced inhibition.



(a) (b)

Figure 3.35: KL divergence between (a) first and subsequent phases and (b) between successive recording phases for culture 77. X-label colours reflect the type of recording phase, corresponding to the colours in Table. 2.1

Chip 184 was observed over a much longer time period (3 weeks) and as a result higher divergences are seen between phases, when computed between the full model and the independent model, Fig. 3.36. However, this is not the case for the H0 model, indicating that while firing rate may fluctuate over periods of 3-7 days, leading to a higher observed divergence in the original models, correlation structure remains more constant, in line with the chip 136 experiment.



(a) (b)

Figure 3.36: KL divergence between (a) first and subsequent phases and (b) between successive recording phases for culture 184

3.5 Discussion

In this chapter, a novel modification to the Ising model was presented and explored in a simulated test bed environment, as well as being applied to a collection of hippocampal culture recordings. This modification involved the rescaling and renormalisation of the full model's pattern distribution, in order to remove, or lessen, the dominant influence of firing rate over these probabilities.

For the controlled simulations, synthetic spike data was generated using a dichotomised Gaussian model so as to allow for the control of individual neural rates and correlation between neurons. In doing so, a perfect test case in which to assess the H0 and αJ models was created. Results indicated the H0 model's ability to uncover correlation changes within the data that were missed by the full Ising model, due to the masking effect of firing rate on the distribution. It was less clear, however, as to the advantages of the αJ model, as while the effect that the rescaling had on the pattern distribution was understood, divergence measures across datasets often mirrored those of the full model, due to the persistence of the bias parameter within the calculation of the pattern distribution.

The H0 model was then applied to a series of *in vitro* hippocampal recordings, in which a selection of the cultures had undergone treatment in the form of CNQX, a

chemical that has been shown to to inhibit network activity by decreasing excitatory synaptic transmission. In accordance with results from the full model, the H0 model identified a higher divergence between recording sessions within those cultures that had undergone induced inhibition, strengthening evidence that an observed permanent remodelling effect had taken place within these networks.

Results from the eigendecomposition of the FIM (full model) also reinforced the concept of sloppiness in cultured neuronal networks [41], and observed differences between stiff eigenvectors before and after CNQX treatment was another strong indication of network remodelling, one that will be pursued further in the following chapter, in which the FIM analysis is extended to larger networks using a diagonal approximation.

While initial results of the uses of the H0 model in relation to simulated networks were promising, more real-life experiments are vital in consolidating the effectiveness of the Ising model renormalisation in isolating changes in correlation structure (see chapter 5 for a further example).

Chapter 4

Large-Scale Analysis for Neuronal Networks

4.1 Introduction

One of the main downsides of exact pairwise models is, of course, their size limitations. It is possible to sample over many groups of neurons within a network as shown in the previous chapter, but even with this technique there are external influences on each group that are not accounted for. When investigating neuronal networks it is important to monitor large-scale activity, and while large-scale models do sacrifice some goodness of fit they offer population wide measures of activity that, in conjunction with smaller more precise models, give a more complete picture of the underlying mechanisms driving network behaviour.

One such model that will be explored in the following chapter is the population tracking model, outlined in chapter 2. The population tracking model has been shown to fit networks of up to 1000 neurons while still allowing for a low dimensional estimation of model of the full pattern probability distributions [32], and has been widely accepted as an accurate model of population activity [7, 34, 35, 36]. These entropy estimations have been used to characterise neural representation of stimuli within the mouse neocortex across development [32] and quantify differences in neural circuit changes in brain disorders [33]. In addition to this model, it is possible to leverage

one of the properties of the pairwise maximum entropy models, namely the FIM, at a much larger scale via a diagonal approximation. As outlined in chapter 2, the Fisher information matrix offers a highly informative analysis of model parameter sensitivity and has been shown to be obtainable directly from spiking activity, provided that sufficient data is available. Using the easily computable diagonal matrix does come with a significant loss of information but, depending on the particular structure of the model in question, it has been shown to be a satisfactory approximation [44, 66, 67].

In the following chapter, these methods of analysis were applied to larger populations of the previously explored cultured datasets in order to further characterise network behaviour across recording phases.

4.2 Results - Population Tracking Model

As seen in previous sections, due to the combinatorial explosion of possible patterns the exact Ising model is limited to a small number of units (N =10-20). Therefore it is of interest to examine models capable of fitting a larger ensemble of neurons. While the population tracking model (Sect. 2.5) can handle larger networks, it still has limitations and so cannot capture the entire collection of recorded neurons captured by the MEAs. Therefore, groups of $180-190$ neurons were selected for fitting. These groups were obtained by pooling together all neurons from the 20 smaller groups (N =10) used in the previous chapter. Duplicated neurons (neurons contained in more than one of smaller groups) were only included once, so if a group contained a higher number of duplicates (and so had a smaller population size), then additional neurons were added by randomly sampling them from the population using the same rate threshold ($0.1-10Hz$).

4.2.1 Model Fits & Evaluation

Model fits can be visualised by plotting the pattern probabilities predicted by the model versus those observed in the data, Fig. 4.1. Here it is clear that overall probabilities are predicted well for patterns that are observed more frequently in the data, but as the patterns become less frequent the uncertainty of model estimates grow. This is due to the fact that low probability patterns cannot be estimated well empirically.

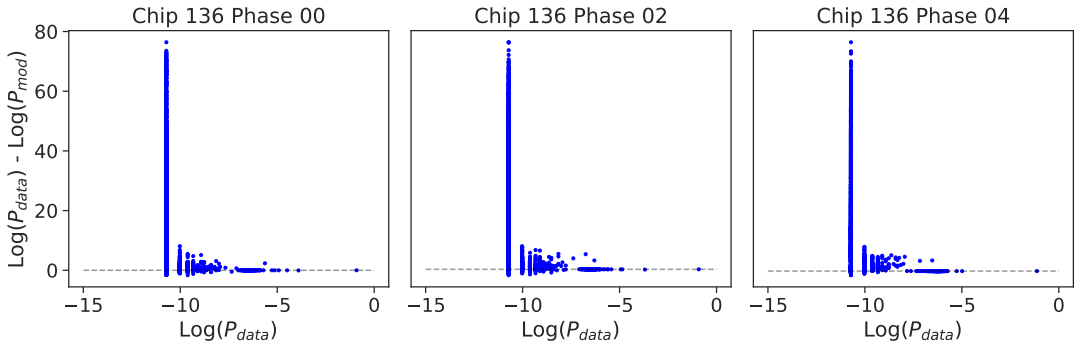


Figure 4.1: Predicted model probabilities versus observed probabilities for the population tracking model.

Taking a zoomed in view of these probabilities while also calculating probabilities predicted from the independent model (i.e. using only neuron rates) exemplifies the superior accuracy of the population tracking model, Fig. 4.2.

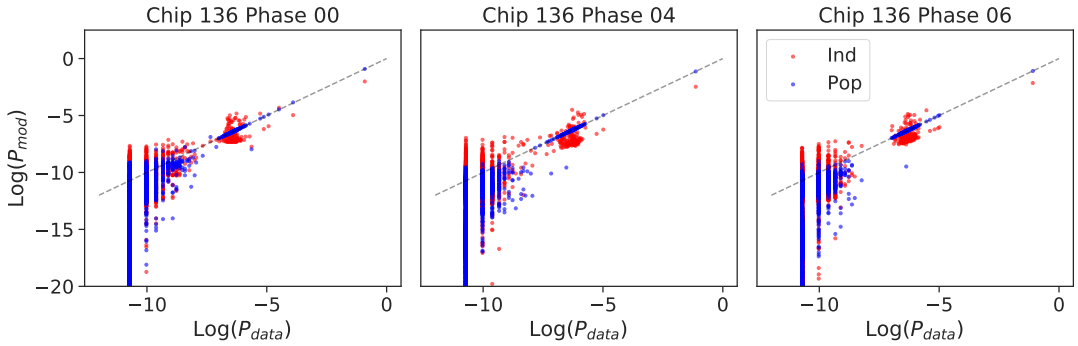


Figure 4.2: Predicted model probabilities versus observed probabilities for both independent and population tracking model (zoomed in to better demonstrate the advantages of the population tracking model).

This can be quantified further by computing the log-likelihood ratio ($LLR = \log(P_{Model}) - \log(P_{Ind})$) as a measure of the model's advantage over the independent model.

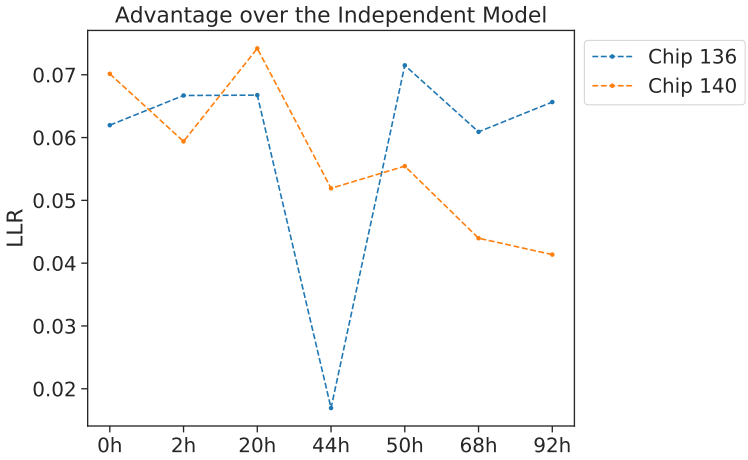


Figure 4.3: LLR of population tracking model across phases for chip 136 and 140

4.2.2 Entropy Estimation

As mentioned previously, the entropy of a distribution can be thought of as a measure of uniformity or lack of structure, distributions in which all patterns have probabilities around the same value (i.e. are similarly likely to occur) will have a higher entropy, whereas distributions in which a small number of patterns are much more likely to occur than the rest (i.e. a sparse distribution) will have a lower entropy. In order to understand this relation, entropy was computed for the population tracking model (H_{full}) and a number of other simplistic models:

1. Homogeneous Model: Each neuron has identical firing, obtained from the mean firing rate of all neurons (r_m):

$$H_{iid} = -N[r_m \log_2(r_m) + (1 - r_m)\log_2(1 - r_m)] \tag{4.1}$$

2. Independent model: Each neuron is assumed independent with rates equal to those observed from the data:

$$H_{ind} = -\sum_{i=1}^{N}[p(x_i)\log_2(p(x_i)) + (1 - p(x_i))\log_2(1 - p(x_i))] \tag{4.2}$$

3. Population Model: Each neuron is identical but with the synchrony distribution $p(k)$ matched to the data:

$$H_{pop} = -\sum_{k=0}^{N}\left[p(k)\log_2(p(k)) + p(k)\log_2\left(\binom{N}{k}\right)\right] \tag{4.3}$$

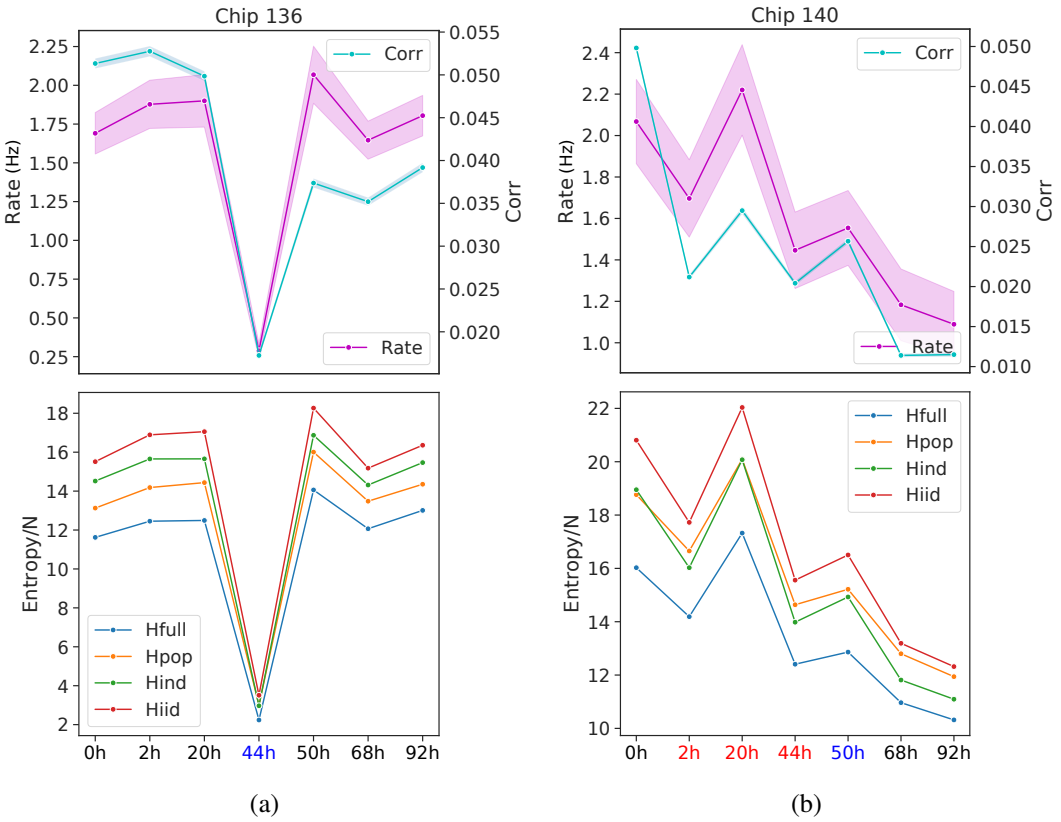
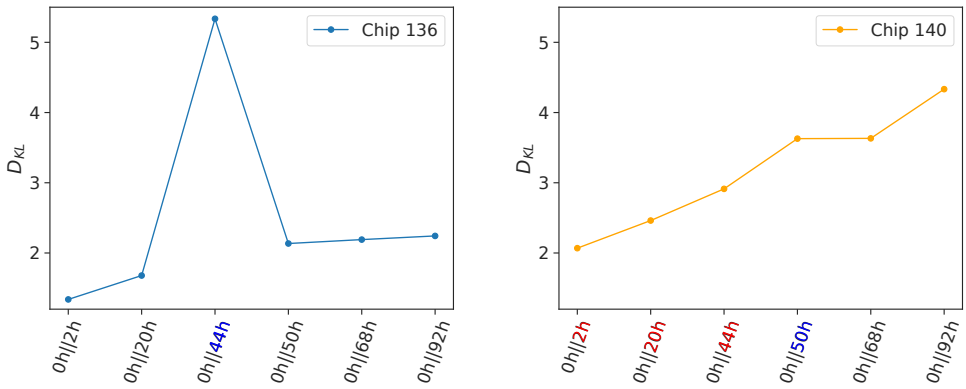


Figure 4.4: Entropy computed using different models of the network population across phases

It is now of interest to understand how network statistics such as firing rate and correlation will influence the shape of the pattern distribution and therefore the entropy. For instance, if the overall firing rate of a population increases, this will lead to a higher probability of spikes occurring, increasing the number of patterns that are observed and hence the entropy. Conversely, if the correlation of a population increases, groups of neurons will spike or remain silent together, leading to a decrease in entropy [32]. These relations can be demonstrated in Figs. 4.4, 4.7, 4.9, where a clear connection between entropy, rate and correlation is displayed. Furthermore, this effect is highlighted by comparing the entropies H_{full} and H_{ind}: while H_{ind} follows the change in mean rate almost exactly, it over estimates the true entropy as it does not account for the reduction in entropy that arises from network correlations, which are incorporated into the calculation of H_{full} through the population synchrony distribution. Fig.

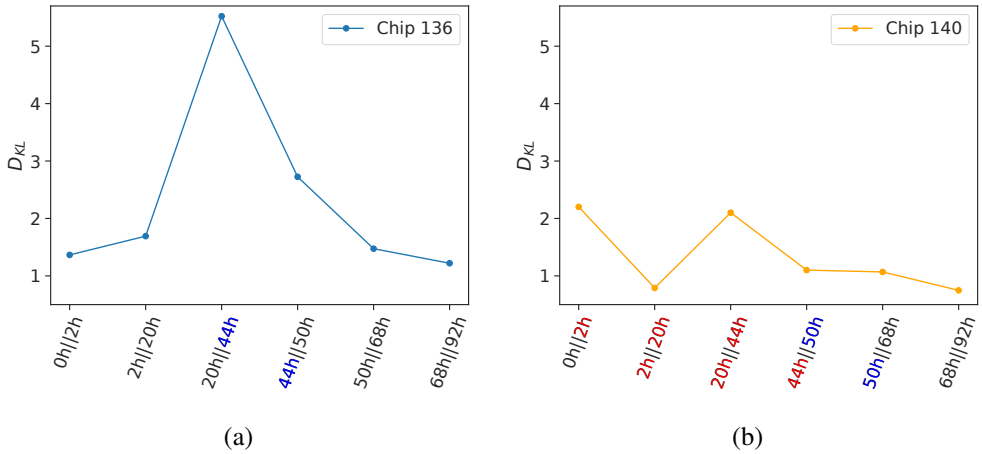
4.4 shows the mean rate and correlation for the selected network groups as well as the entropies computed from each model across recording phases, for cultures 136 and 140. In terms of activity, both cultures initially display similar levels before the CNQX addition inhibits activity in chip 140. After washout, chip 136 recovers activity levels whereas 140 remains low, indicating that a remodelling effect has occurred within the network. Entropy levels further reflect this balance between rate and correlation across phases, demonstrating the reduction in information coding capabilities incurred via induced inhibition. KL divergence was also computed using the full model between the first recording and the subsequent recording phases throughout the experiment, Fig. 4.5. The higher divergence obtained for chip 140 indicates a higher difference between resulting distributions across time, reinforcing evidence of a remodelling effect in this culture.



(a) (b)

Figure 4.5: KL divergence computed from the population tracking model between original and subsequent datasets across time

Additionally, Fig. 4.6 again shows the divergence computed from the model, only this time between successive recordings in the experiment. Results indicate a lower divergence observed within the network that underwent remodelling (140) once the CNQX had take effect, i.e. after remodelling had taken place.



(a) (b)

Figure 4.6: KL divergence computed from the population tracking model between successive datasets

Chip 77 received two separate dosages of CNQX, one of 5μM and one of 10μM, that were monitored over a shorter time frame compared to chip 140 (24 hours compared to 48). Initially chip 77 exhibited lower levels of activity compared to the previous two cultures, but a sharp rise is displayed after the wash, indicating that a full remodelling effect is not observed during the shortened 24 hour time frame. The second, stronger dosage of CNQX then ensures that activity remains low.

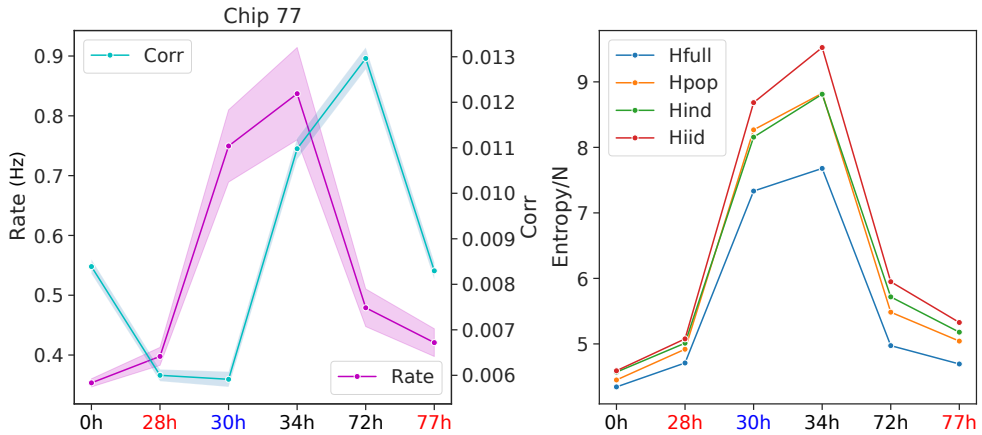


Figure 4.7: Entropy computed using different models of the network population across phases

KL divergence between successive recording sessions suggest a relatively high stability across phases, with the largest change coming between the two recordings after the first CNQX addition and washout.

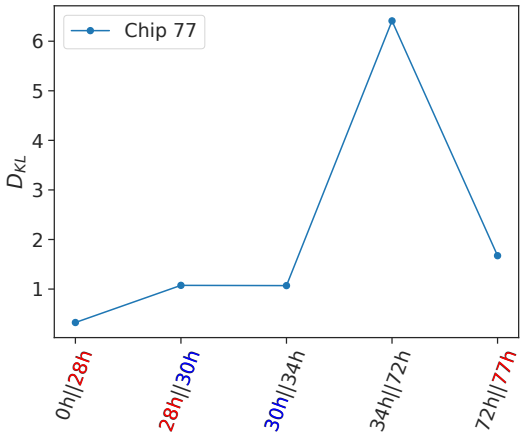


Figure 4.8: KL divergence computed from the population tracking model between successive datasets

Chip 184 was recorded over a longer period (3 weeks) without undergoing any CNQX treatment, and displayed a high stability which is reflected in the low divergences between models, even across periods of up to 7 days, Fig. 4.10.

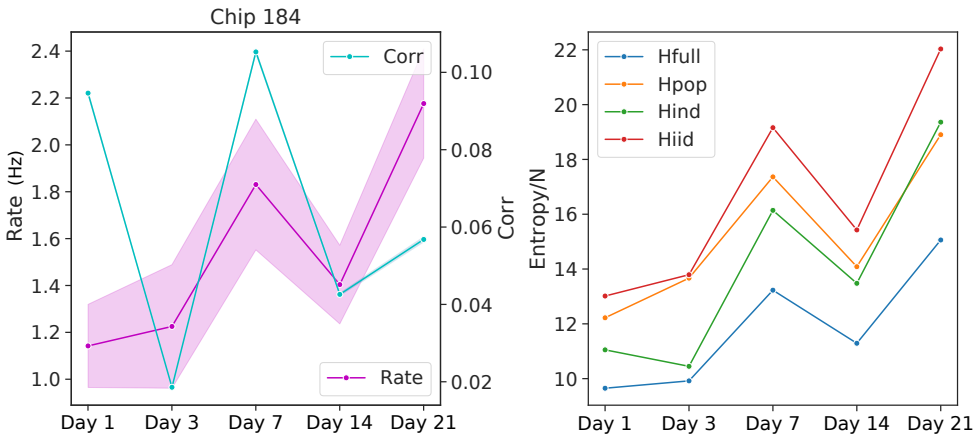


Figure 4.9: Entropy computed using different models of the network population across phases

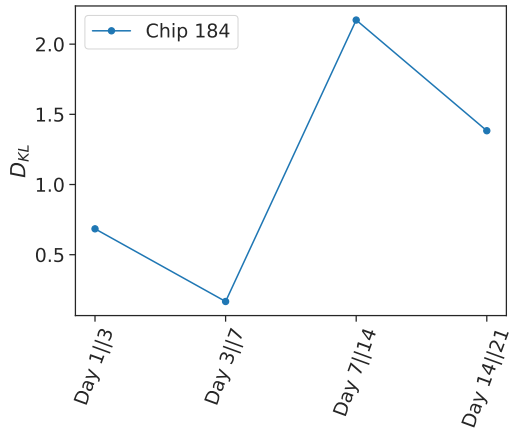


Figure 4.10: KL divergence computed from the population tracking model between successive datasets

Fig. 4.11 shows the average KL divergence measures computed per chip, highlighting the overall differences in divergence values for each experiment.

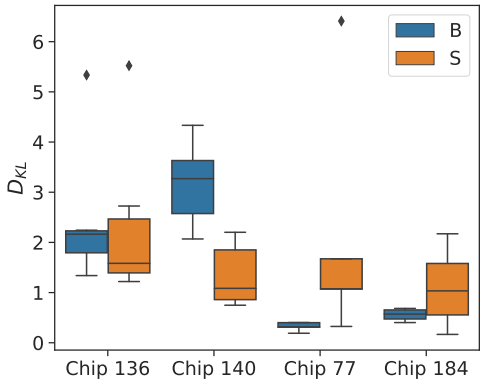


Figure 4.11: Boxplot demonstrating the different KL divergence values per chip. B indicates measures carried out with the baseline recording and S indicates measures carried out between successive recordings

4.3 Results - The Fisher Information Approximation

As discussed in chapter 2, the Fisher Information Matrix (FIM) offers a quantification of parameter sensitivity within a model. As a reminder, the entries of the FIM for the Ising model are computed as follows

$$f_{lm}(\Theta) = < X_l X_m >_{model} - < X_l >< X_m >_{model} \tag{4.4}$$

where $X_l = x_j$ for the parameters λ_j and $X_m = x_{jk}$ for the pairwise parameters λ_{jk}. This can also be approximated from the data itself by taking expectations with respect to the probability distribution $P_{data}(\mathbf{x})$. These two versions of the FIM have been shown to be almost identical which constitutes the first simplifying step of the Fisher approximation. However, even without fitting a model, calculating this version of the FIM directly from data still requires $N_F(N_F+1)/2$ terms to be computed and so the combinatorial explosion of entries still pertains. This leads to the second step of the approximation, which requires that only the diagonal entries need to be calculated:

$$F' = \begin{pmatrix} f_{h_1,h_1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & f_{h_N,h_N} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & f_{J_{1,2},J_{1,2}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & f_{J_{N-1,N},J_{N-1,N}} \end{pmatrix} \tag{4.5}$$

Recalling that the FIM defines a metric in parameter space, the full FIM contains vectors in each direction (forming an ellipse as in Fig. 2.3) while the diagonal approximation can be interpreted as the intersection between this ellipse and the parameter axis, Fig. 4.12. While this comes with a great loss of information, recent work has shown that the diagonal of the FIM is a satisfactory approximation of the full matrix when dealing with large networks [66, 67], provided that sufficient data is available.

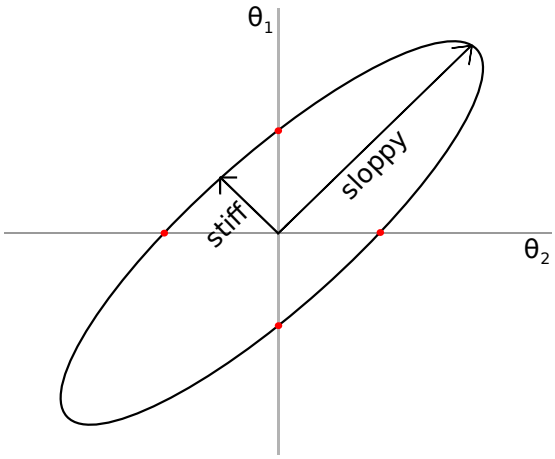


Figure 4.12: Schematic demonstrating an ellipse determined by the full FIM w.r.t two parameters and the intersection of this ellipse with the parameter axis (points in red) detailing the diagonal approximation

Given this interpretation of the diagonal FIM, it is now of interest to examine what this looks like for neuronal activity data. This can be investigated by considering an example from the previous chapter in which the full FIM was calculated for a number of smaller subgroups. Here, it is possible to compare the eigendecomposition of the FIM to the diagonal approximation, Fig. 4.13, by reshaping the diagonal into an $N \times N$ matrix. Interestingly, the two matrices display a highly similar structure, therefore offering strong evidence for the validity of the approximation. Now, the FIM calculation is far more efficient, requiring only $N_F = N(N+1)/2$ computations while still offering an informative view of the population activity.

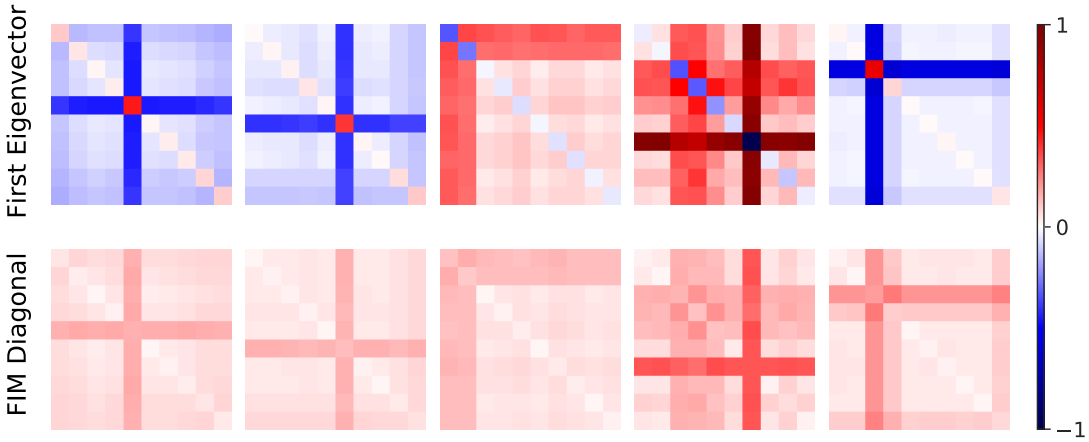


Figure 4.13: Demonstration of the similarities between the first eigenvector of the full FIM and the reshaped diagonal entries, for five different sample subpopulation groups (N=10) taken from the Ising model experiments. Structure is highly similar between the two.

4.3.1 Parameter Structure

To give an idea of the form of this approximation Fig. 4.14 shows the structure of the diagonal elements of the FIM computed from data obtained from three consecutive chip 136 recordings. Clearly, the structure of these three matrices is highly conserved, suggesting that model parameters remain stable over many hours of recording phases. This supports the findings of previous studies that examined the stability of small sub-networks within a set of neuronal cultures by fitting Ising models and carrying out FIM analysis of the models across time [41, 96]. As the FIM is already diagonal, there is no room for such eigendecomposition analysis as the entries along the diagonal already define the matrix eigenvalues. As a comparison, these entries are normalised and ranked to exemplify the difference in structure to the full model, Fig. 4.15.

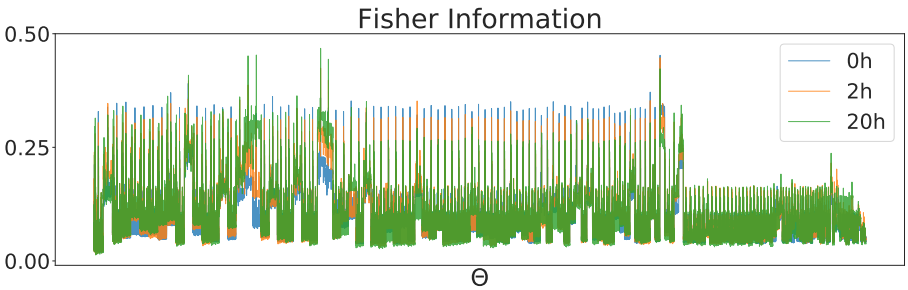


Figure 4.14: Diagonal entries of the FIM computed for three recording phases (chip 136)

Here, the matrix exhibits a far more uniform rank of values, with 80% variance being achieved at around $\sim 70\%$ of parameters. In a similar vein to the analysis outlined in Sect. 2.4.2, where eigenvectors were reshaped in order to correspond more intuitively to model parameters, the diagonal FIM can be converted into an $N \times N$ matrix, as can be seen in Fig. 4.16 (chip 136).

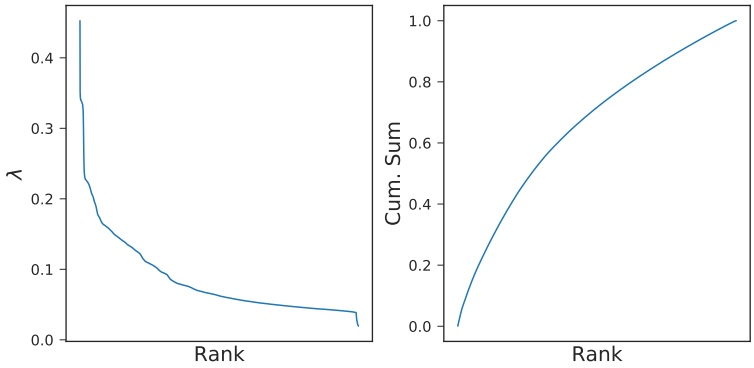


Figure 4.15: Ranked diagonal entries (right) and cumulative sum (left)

As shown in the previous section, the form of this matrix has similarities to the principal eigenvectors of the full model, highlighting a clear structure of higher sensitivities belonging to a small subnetwork of neurons within the population (or the parameters belonging to these neurons). Although this is not a direct quantification of sloppiness at a larger scale within neuronal networks, it certainly points to the possibility of a similar network mechanism exhibited at the population level. Additionally, as has been observed in prior analysis, this structure is highly conserved across the course of the experiment, again indicating a strong stability within the cultured networks.

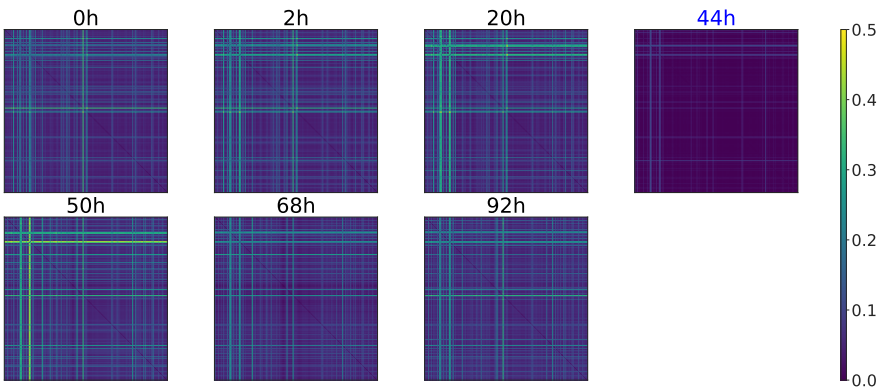


Figure 4.16: Chip 136: the diagonal FIM is reshaped into an $N \times N$ matrix and plotted across recording phases

Chip 140 also displays a strong stability across phases, although key differences can be observed before and after the CNQX effect takes place. As time passes, the matrix displays a sparser structure compared to its original state, indicating that while the remodelling effect inhibits activity and correlation in the network, the 'stiff' neurons within the network are robust to change.

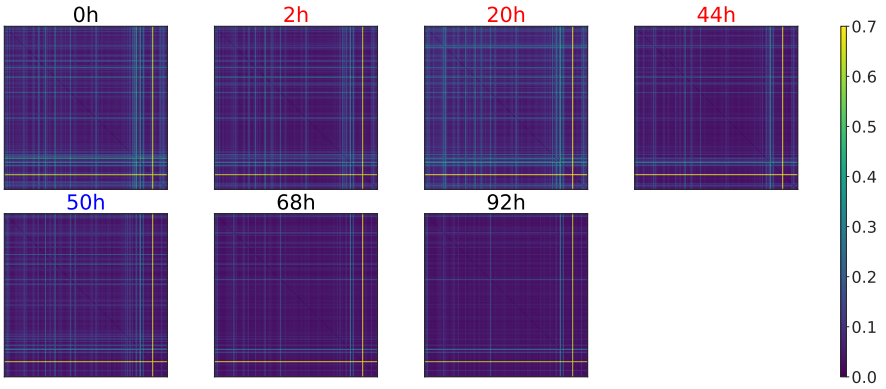


Figure 4.17: Chip 140: the diagonal FIM is reshaped into an $N \times N$ matrix and plotted across recording phases

Chip 77 exhibited less stable behaviour, with changes in parameter structure observed during and after the addition of CNQX. Although this may be down in part to the variable nature of the activity across the recording phases.

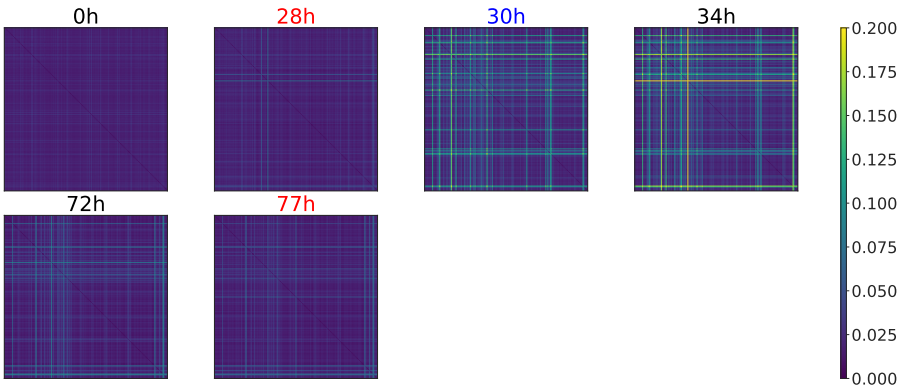


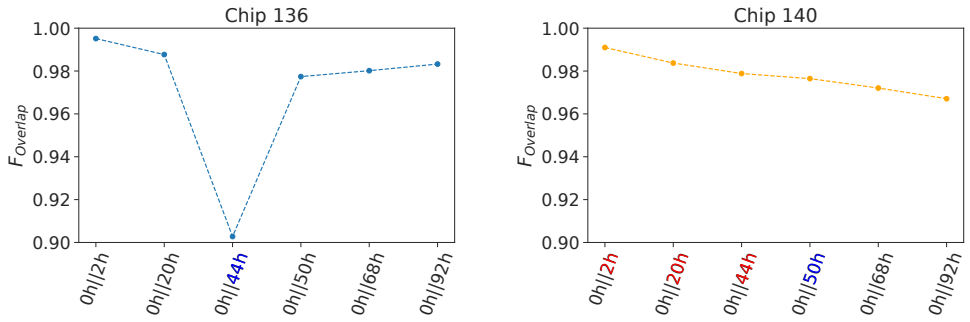
Figure 4.18: Chip 77: the diagonal FIM is reshaped into an $N \times N$ matrix and plotted across recording phases

Chip 184, a culture that was observed over a long period (3 weeks) in the absence of external inhibition, remained extremely stable across the entire duration, Fig. 4.19.

Figure 4.19: Chip 184: the diagonal FIM is reshaped into an $N \times N$ matrix and plotted across recording phases

4.3.2 Fisher Overlap

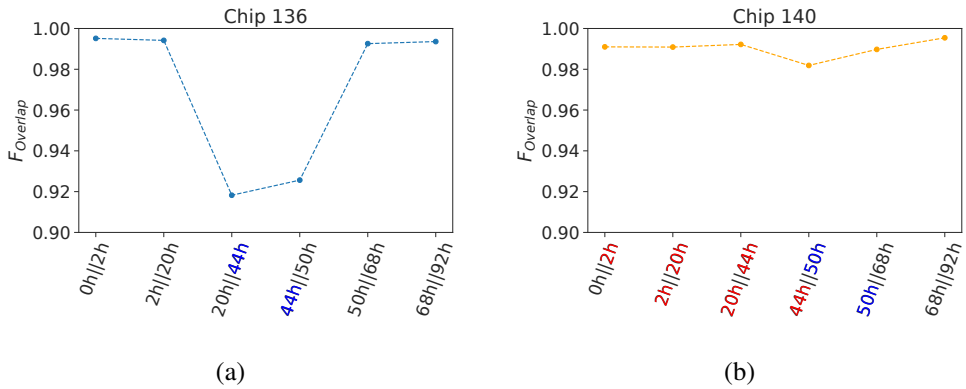
In order to assess the similarity between datasets across the experiment the Fisher overlap was utilised (see Sect. 2.4.4).



(a) (b)

Figure 4.20: Fisher overlap computed between original and subsequent datasets across time

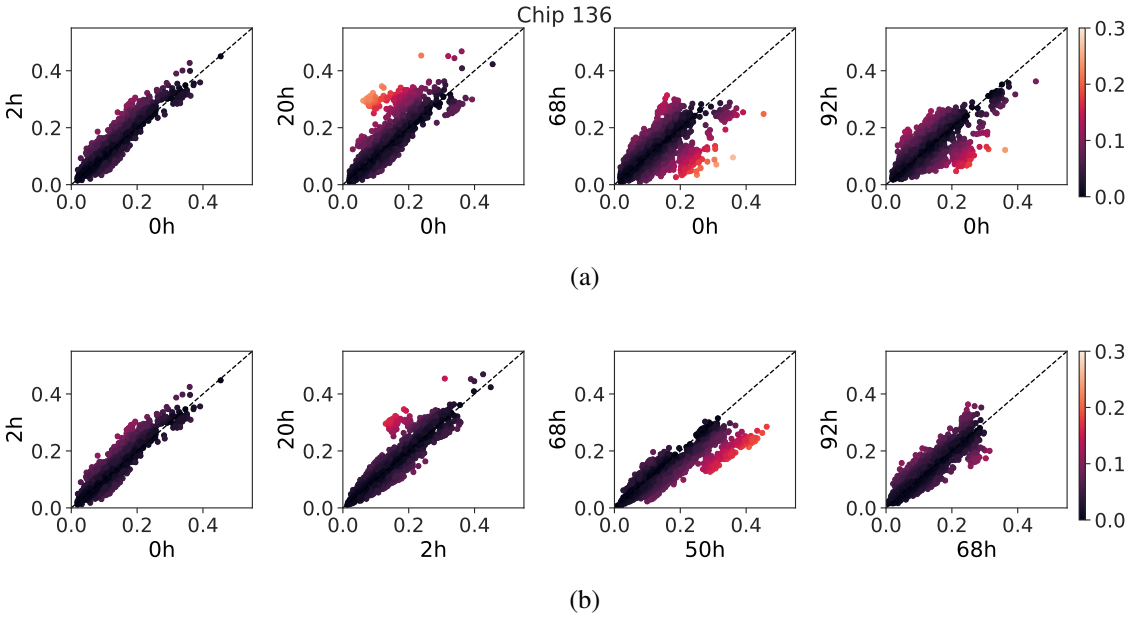
Fig. 4.20 shows the Fisher overlap computed between the original and subsequent recordings for chips 136 and 140. Both cultures exhibit a high degree of overlap, suggesting that the ordering parameter structure remains relatively stable across the course of the experiment. Further to this, chip 140 displays a lower overlap with the original recording compared to 136, which becomes lower still after the CNQX effect takes place. However, the overlap between successive datasets is higher in 140 after this effect, Fig. 4.21, indicating a higher similarity in network structure after remodelling.



(a) (b)

Figure 4.21: Fisher overlap computed between successive datasets across time

To understand better where this difference in overlap is coming from, individual FIM entries were compared across recordings both in terms of difference from the original recording and between successive recordings. High correlation between FIM values across recordings indicates that chip 136 (control experiment) maintains a stable structure in terms of parameter importance ordering, Fig. 4.22, and while some changes are observed after washout, high similarity in structure observed between the first and last recording give a further indication of a robust circuitry across phases.



(a)

(b)

Figure 4.22: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording. Colour map represents the absolute values of the difference between the two FIMs, ($F_d = |F_1 - F_2|$)

However, in the chip 140 experiment, there is a clear subgroup of parameters at 0 hours that undergoes a significant shift during the CNQX phase and remain changed after washout. This isolated cluster of parameters is visible in all comparisons to the baseline recording, Fig. 4.23(a), but strong correlation between successive phases thereafter, Fig. 4.23(b), indicates a permanent circuit reorganisation within the network.

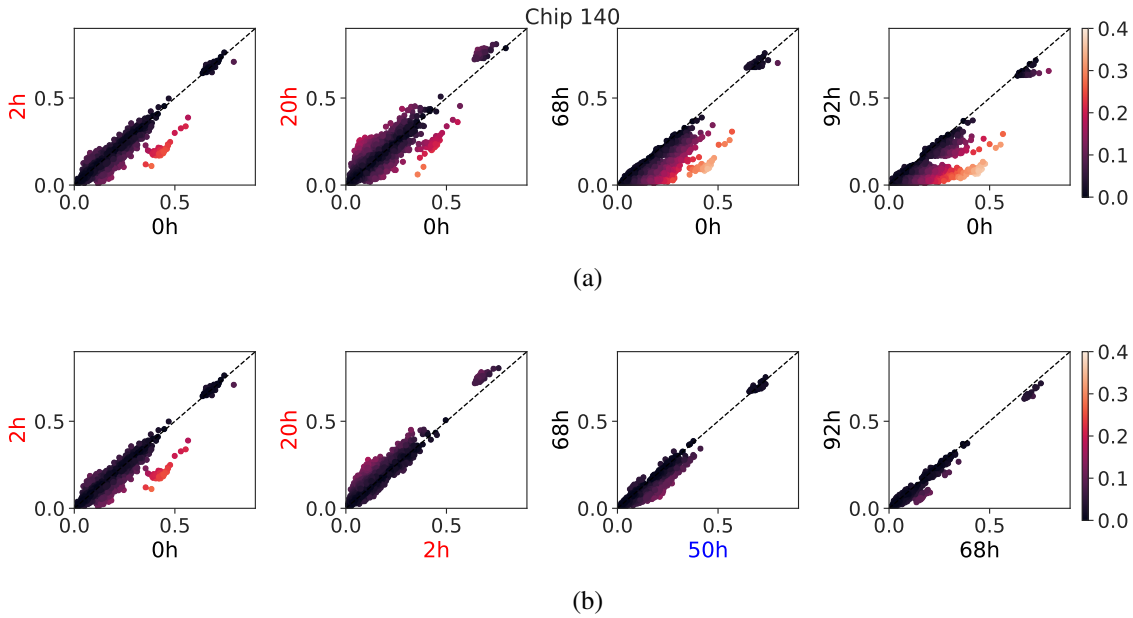
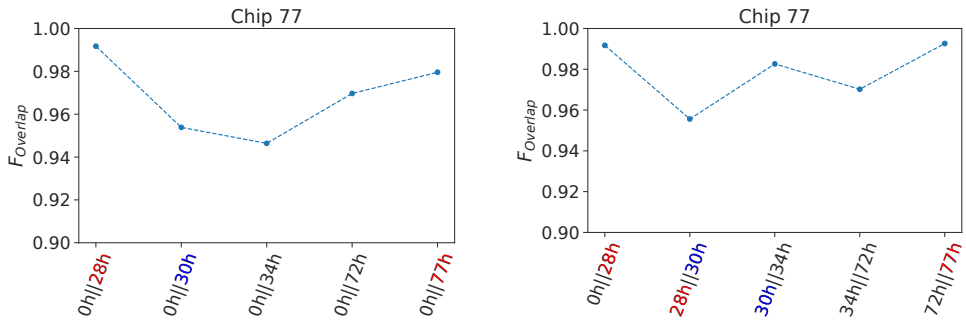


Figure 4.23: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording

As can be expected from Fig. 4.18, chip 77 displays a lower overlap between phases in which the diagonal FIM is significantly changed, a clear sign of the variable nature of the network as a result of multiple CNQX additions.



(a) (b)

Figure 4.24: Fisher overlap computed between (a) original and subsequent datasets and (b) successive datasets across time

Additionally, scatter plots indicate significant shifts in parameter sensitivity after both periods of CNQX addition, Fig. 4.25

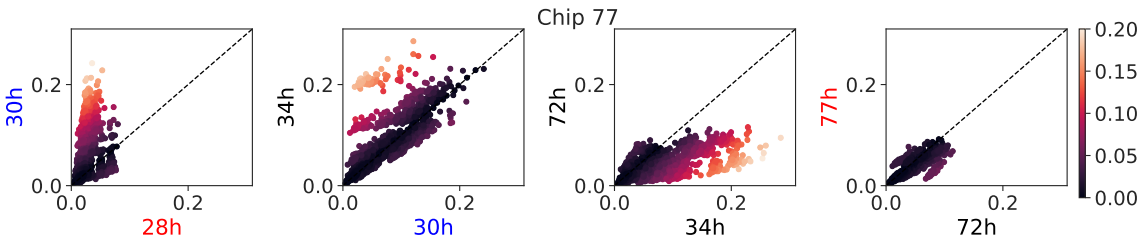
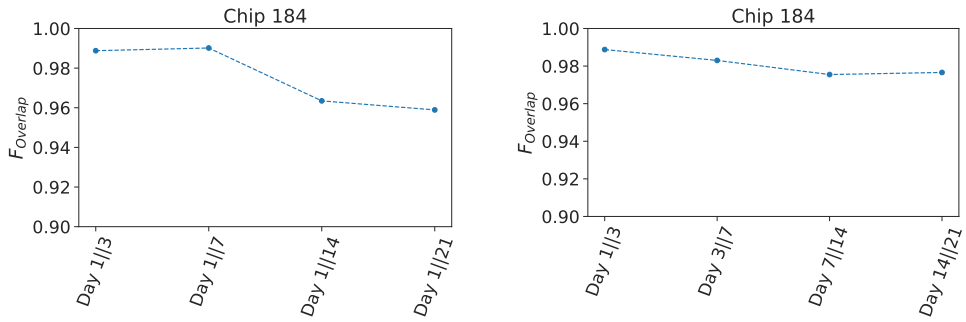


Figure 4.25: Scatter plots detailing change in FIM values compared to the previous recording

Chip 184 was found to have a high overlap throughout the whole experimental period, with similarity becoming smaller as the time between recording sessions grew. This is again supported by a strong and consistent correlation in FIM diagonal across all phases, Fig. 4.28.



(a) (b)

Figure 4.26: Fisher overlap computed between (a) original and subsequent datasets and (b) successive datasets across time

As a comparison, the Fisher overlap was computed using the simulated DG datasets from the previous chapter, as can be seen in Fig. 4.27. Here, values were obtained by taking the diagonal Fisher matrices of all groups from the baseline network and calculating the overlap between these values and the corresponding values as the generating parameters were increased. Results indicated that overall, Fisher overlap was comparable for the simulated and recorded datasets.

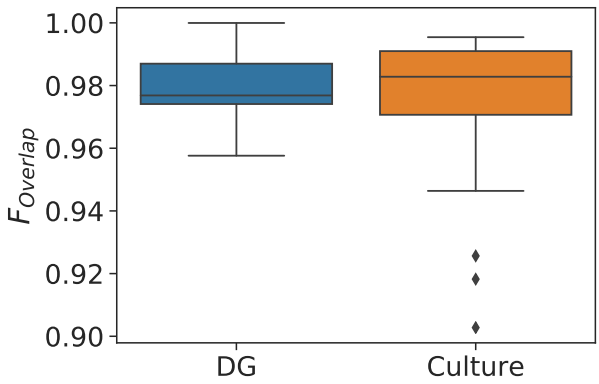


Figure 4.27: Fisher overlap values obtained from the simulated (DG), and recorded (Culture) datasets.

Another interesting feature of the FIM scatter plots is the possibility to identify individual neurons over time. This can be seen in Figs. 4.29-4.32, where the FIM entries corresponding to each neuron's bias parameter (h_i) are colour coded. The effect is an observed clustering of parameters belonging to single neurons (the h_i's plus the

J_{ij}'s), albeit only for the highly sensitive, 'stiff' parameters ('sloppy' parameters tend to be indistinguishable and bunched around zero). This uncovers some extra information about the observed circuit changes by identifying which neurons are predominantly involved.

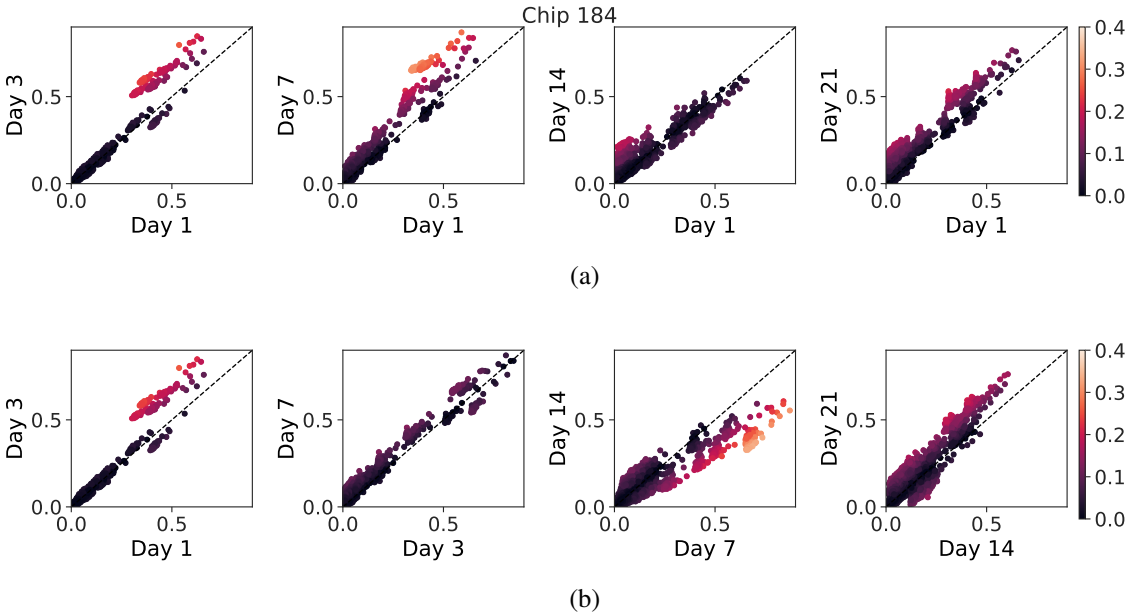


Figure 4.28: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording

For example, the original shift in parameter sensitivities observed in chip 140 (from 0-2 hours) primarily belong to a single neuron, with more neurons becoming involved later in the experiment, Fig. 4.30. A number of neurons can also be seen undergoing significant change in sensitivities during the two CNQX phases of the chip 77 experiment, Fig. 4.31.

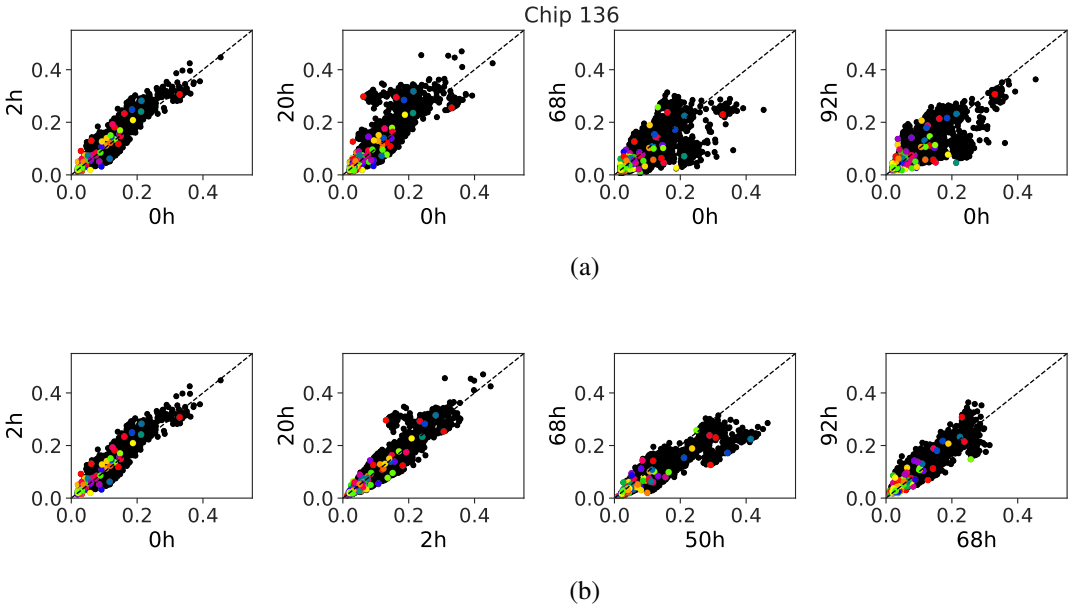


Figure 4.29: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording. Coloured dots represent the h_i parameters

Chip 184, the most stable of the four experiments, exhibits a number of distinct clusters, Fig. 4.32. While these cluster do undergo shifts, the ordering of parameter sensitivities remains relatively stable over the experimental period.

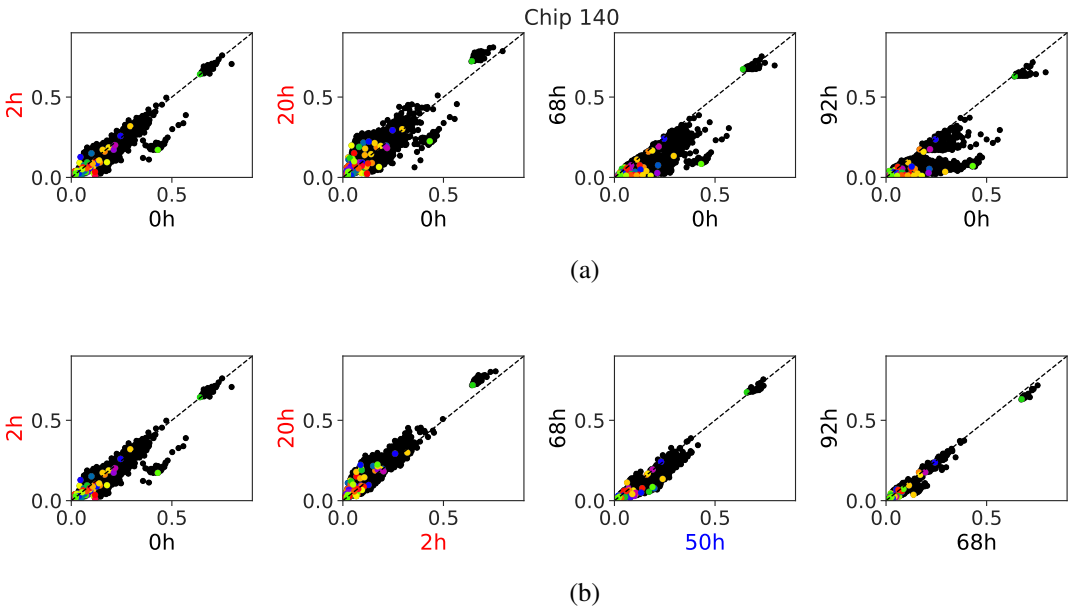


Figure 4.30: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording

Overall, the Fisher approximation provided some key insights into network function across the experiment by offering supporting evidence for the stability of key neurons across time and also highlighting differences between networks in the presence and absence of CNQX.

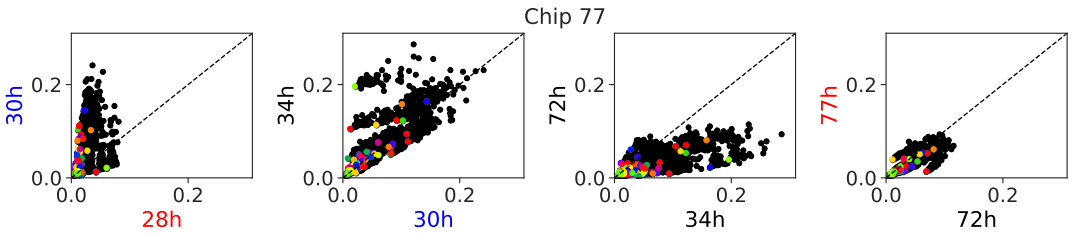


Figure 4.31: Scatter plots detailing change in FIM values compared to the previous recording

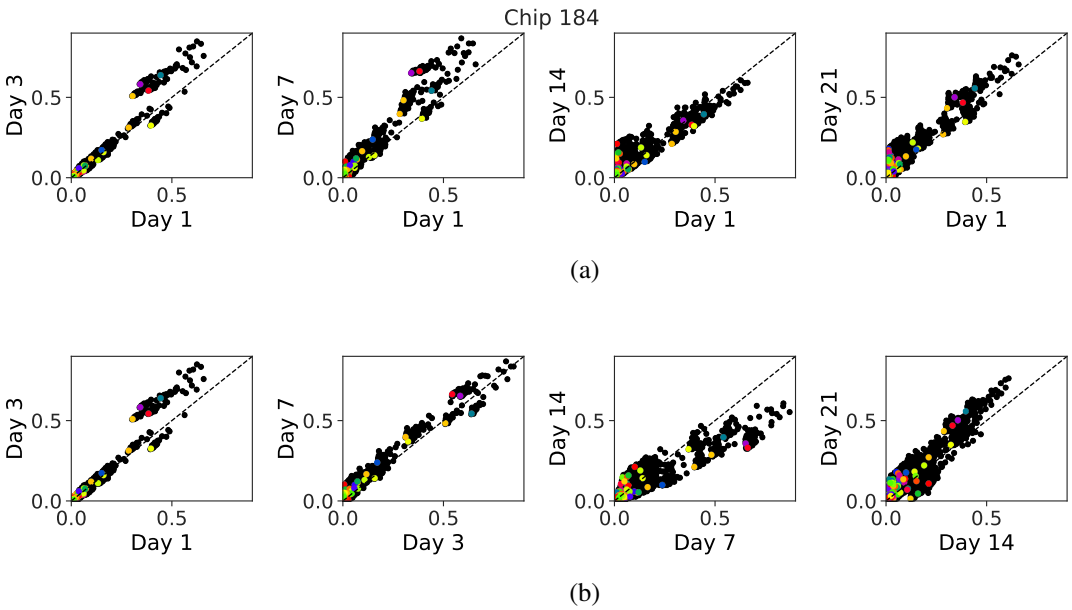


Figure 4.32: Scatter plots detailing change in FIM values (a) compared to baseline recording and (b) compared to the previous recording

4.4 Discussion

The large scale analysis of cultured network activity presented in this chapter further strengthens claims of a network wide remodelling effect as a result of CNQX treatment. Firstly, entropy measures offered a characterisation of population wide activity across experiments while KL divergence further quantified the differences between

pattern probability distributions before and after the CNQX effect had taken place. This was exemplified by the higher divergences observed in CNQX experiments before and after application and the lower divergences in successive recordings after the remodelling had taken place.

The diagonal Fisher information was shown to be a reasonable approximation of the full model for smaller networks and its extension to larger populations again highlighted the observed change in network organisation via the computation of the Fisher overlap, with results here aligning with the KL divergence tests. Furthermore, direct comparison of the FIM diagonal at different points in the experiment enabled interesting insights into parameter structure, uncovering an efficient method for discovering both stability levels and organisational shifts in parameter sensitivities across recording phases. Additionally, observed clustering in parameter sensitivities allowed for the identification of individual neurons within the FIM structure, enabling a novel and computationally inexpensive method of identifying important (or sensitive) neurons in large neuronal populations.

The FIM analysis was of particular interest due to the high similarity observed between the 'stiff' directions of the full model and the approximated diagonal FIM, leading to the exciting possibility that characteristic sloppy model mechanisms, already found on a smaller scale in [41], could exist at the population level. Such a finding would offer a fascinating insight to network behaviour and is an exciting avenue for future work.

While all results here offer strong evidence of a circuit reorganisation, more data would be vital in providing a more concrete proof of this conclusion.

Chapter 5

Characterising Neural Population Coding in SynGAP$^{+/-}$ Mice

5.1 Introduction

Autism spectrum disorders (ASD) are a broad set of neurodevelopment disorders characterized by impairments in social interactions and communication as well as restricted, repetitive, and stereotyped patterns of behaviour that typically emerge in the first few years of life [97]. ASD commonly coincides with instances of intellectual disability (ID), and together are the most common developmental disorders present in humans. Combined, they affect between 3-5% of the population. The presence of an intellectual disability, communication disorder, epilepsy or other complicating genetic disorders occurs in approximately 70% of people with ASD [98].

ID and ASD can arise from a diverse array of genetic and environmental factors, with a small fraction of these stemming from mutations to a single gene (monogenic). From these genetic links, the vast majority involve genes regulating neural activity and plasticity. Studies on the associated monogenic disorders in mice have uncovered differences in neuronal properties at the cellular level (e.g. excitability, synaptic plasticity and synaptic transition [99, 100]) as well as behavioural irregularities but relatively little is known about how these differences propagate to the network and circuit level. In

this study, one of the two most common monogenic disorders: SYNGAP1 haploinsufficiency [101], is investigated using SynGAP$^{+/-}$ mice as an established mouse model of the human disorder.

The *Syngap* gene regulates the encoding of the synaptic RasGTPase-activating protein 1 (SynGAP), a neuronal RasGTPase activating protein (RasGAP) found in excitatory synapses as part of the post-synaptic density (PSD), Fig. 5.1. SynGAP is expressed solely in the brain (most commonly found in the cortex and hippocampus) and has a known function in N-methyl-D-asparate (NMDA) receptor-mediated synaptic plasticity and insertion of AMPA receptors into the postsynaptic membrane, enabling long term potentiation (LTP) [99, 100].

On the cellular level SynGAP$^{+/-}$ mice have been shown to exhibit accelerated maturation of dendritic spines, leading to larger overall spine size and a higher number of mushroom shaped spines compared to thin spines ($\sim 20\%$) [100]. Studies have shown that thin spines are more transient compared with the more stable mushroom spines and are more likely to undergo LTP [102]. The increase in spine size was also consistent with an increase in membrane bound AMPA receptors and a decrease in silent synapses [100, 103]. These neurons exhibited higher frequencies and larger amplitudes of miniature excitatory postsynaptic potentials (mEPSP) [104].

In order to investigate and characterise circuit level impairments in SynGAP$^{+/-}$ mice neural recording datasets of up to ~ 200 neurons were acquired using two-photon calcium imaging of the primary visual cortex, in both SynGAP$^{+/-}$ (HET) and wild-type (WT) animals. The mice were head fixed but free to move on a treadmill and were recorded in response to phase-reversing stimulus (evoked activity) and in darkness (spontaneous activity). In addition to the initial imaging process the animals then underwent a common monocular deprivation (MD) paradigm [105] in order to induce plasticity and further investigate differences between the two genotypes. The neural activity was compared on both the population and single cell level using Ca^{2+} fluorescence levels and inferred spiking activity. By using spikes, it was possible to compute further population measures such as entropy and divergence to characterise network coding efficiency.

Results on obtained population measures such as mean firing rate, correlation and population coupling were comparable between the two genotypes, with no significant differences being found between animals. Additionally, the full distributions of these

quantities at the single cell level showed no significant differences. Orientation selectivity and proportion of visually responsive cells were again similar between the two animal groups. When analysing the average response of each genotype within stimulus presentation there was a tendency for the HET animals to exhibit a larger response earlier on in the presentation compared to the WT animals, with this characteristic response shape remaining constant throughout trials, suggesting a lack of attenuation observed within the SynGAP$^{+/-}$ animals. Results obtained from the population tracking model revealed a similar level of coding capability in the two genotypes, reflected in the population entropy computed for each neural ensemble. Furthermore, ocular dominance measurements were highly noisy across the MD period and a clear plasticity effect was not observed beyond the inherent response variability.

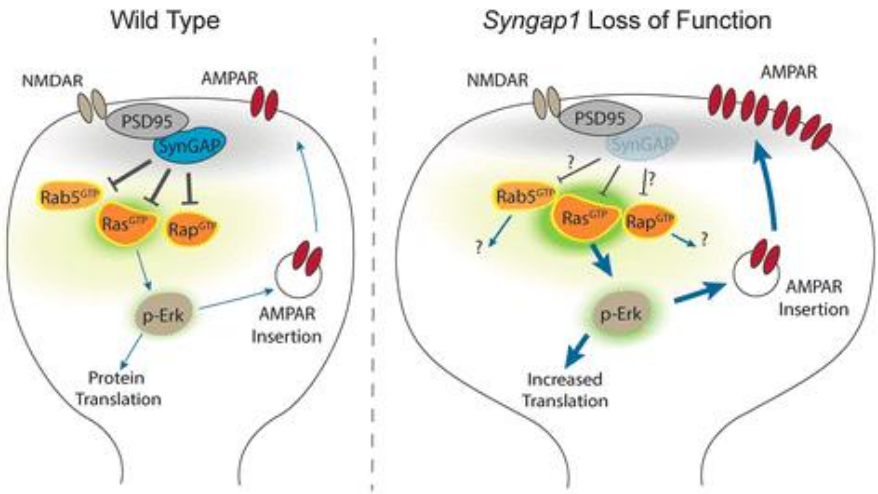


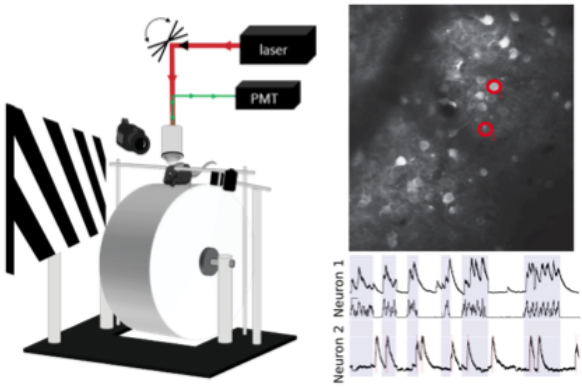
Figure 5.1: Schematic of signalling pathways regulated by SYNGAP1 gene [106]

5.2 Experimental Procedure

5.2.1 Ca^{2+} Imaging

As discussed in Sect. 2.6.3, CA^{2+} imaging is an ideal choice for studying the effects of experience-dependent plasticity as it allows for the direct comparison of individual neurons over time, i.e. before and after plasticity is induced, an advantage that is lacking in other methods such as electrophysiological recordings. Imaging sessions were carried out on awake behaving mice using a resonant scanning two-photon microscope, Fig. 5.2(a). During sessions, changes in somatic fluorescence of a genetically encoded

calcium indicator (GCaMP6f) were recorded over a full field of view, corresponding to an area of around 100 - 200 neurons in primary visual cortex layer 2/3.



(a) (b)

Figure 5.2: (a) Experimental setup of Ca^{2+} imaging for mouse on treadmill (b) Example fluorescence traces of 2 neurons (bottom) extracted from regions of interest (red circles) of Ca^{2+} imaging video frame (top). Grey areas reflect periods of locomotion

Once a recording was obtained, regions of interest (ROIs) from the field of view were drawn (by hand) in order to select areas of the image that corresponded to neurons, Fig. 5.2(b). Mean fluorescence levels of these ROIs, after some signal processing and decontamination [107], were then used as a measure of neuronal spiking activity.

5.2.2 Experimental Paradigm

In order to test the differences in stimulus encoding between the two genotypes a well-established stimulus protocol was used. Each animal session contained four groups: a binocular stimulation to begin with, followed by an ipsilateral stimulation, a contralateral stimulation, and another binocular stimulation to finish (Fig. 5.3). In each group the mouse was subjected to a screen displaying a series of $1Hz$ sinusoidal phase-reversing grating stimuli of varying orientation (0°, 45°, 90°, 135°) with intermittent grey periods.

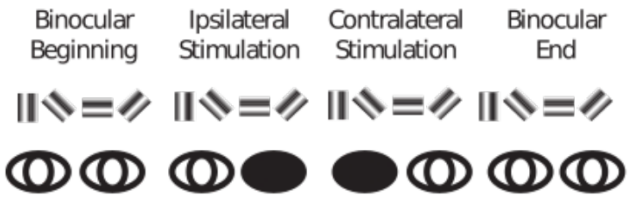


Figure 5.3: Experimental paradigm and stimulus protocol

This standard protocol allowed for a direct comparison of neuronal circuit activity between genotypes, enabling any differences/similarities in network coding capabilities of visual stimuli to be quantified.

5.2.3 Monocular Deprivation

Further to the above experimental method, in which coding capabilities were explored, a secondary aim was to test the differences in cortical plasticity between the two genotypes. In order to investigate this a well-established paradigm for monocular deprivation (MD) was used [105]. The procedure consisted of two identical recording sessions (as described above) before and after a monocular deprivation period, (Fig. 5.4). The idea here was to investigate how the cortical circuits remodelled in response to only receiving input signal from one eye. The animals were divided into two groups, an MD group and a control group. Animals in the MD group had their contralateral eye sutured shut for seven days whereas in the control group the mice were left with binocular vision. This was to ensure that any differences observed across the experiment were significantly different to those observed due to natural fluctuations in neuronal activity across a 7-day period. Having obtained this, the experiment allowed for the direct comparison of individual neuron preference both in regards to orientation selectivity and ocular dominance, before and after MD, shedding light on how the plasticity mechanisms varied across genotype.

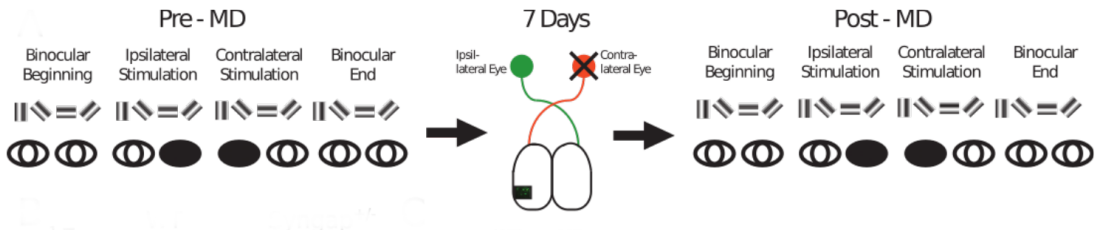


Figure 5.4: Experimental paradigm for monocular deprivation experiment

All experiments were carried out by Evelyn Dylda and Danai Katsanevaki in the laboratory of Nathalie Rochefort at the University of Edinburgh.

5.2.4 Spike Inference

As many statistical models of neural activity model binary neural spike trains, it was first necessary to infer the underlying spike activity modulating the Ca^{2+} fluorescence traces. A number of spike inference software packages exist, the one utilised in this work was MLSpike [108], chosen for its superior accuracy. The algorithm uses maximum likelihood estimation together with a physiological model of intracellular calcium dynamics in order to infer spikes.

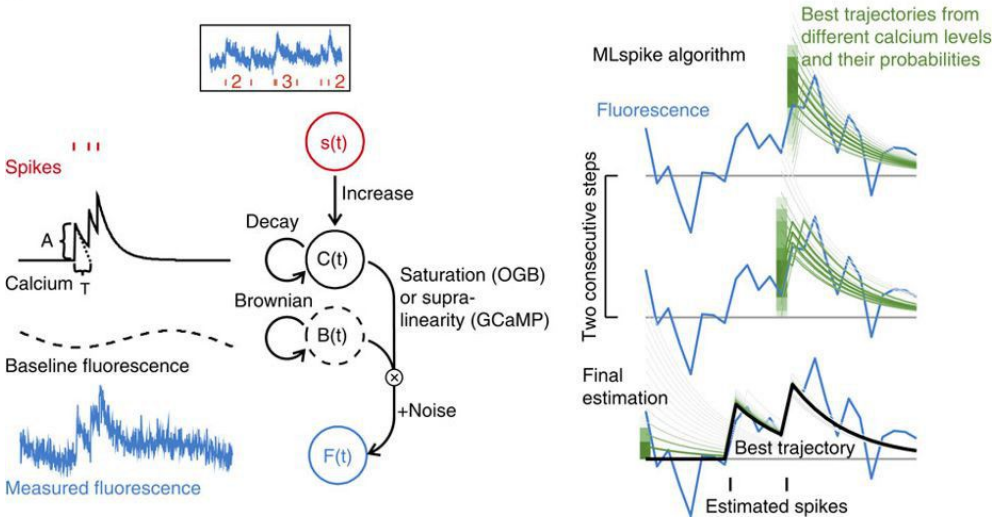


Figure 5.5: MLSpike physiological model and algorithm [108]

The model of intracellular Ca^{2+} concentration C(t) is driven by an amplitude A $\times$ s(t), where A is the unitary calcium response and s(t) is the number of spikes emitted at time t. C(t) then decays to its resting value with a time constant τ. The observed fluorescence F(t) is a product of a drifting baseline function and a nonlinear function of C(t) with some additional noise. The nonlinear function accounts for the dye saturation and GCaMP nonlinearities, which has been extensively modelled to calculate parameters for different indicators. The MLSpike algorithm then calculates the probabilities of the most likely trajectories initiating from all possible Ca^{2+} values, starting at time t and iterating backwards. Once at 0, conditional probability maximisation defines a unique maximum posterior spike train. Fig. 5.6 shows an example spike train inferred from a fluorescence trace recorded from a visually responsive cell.

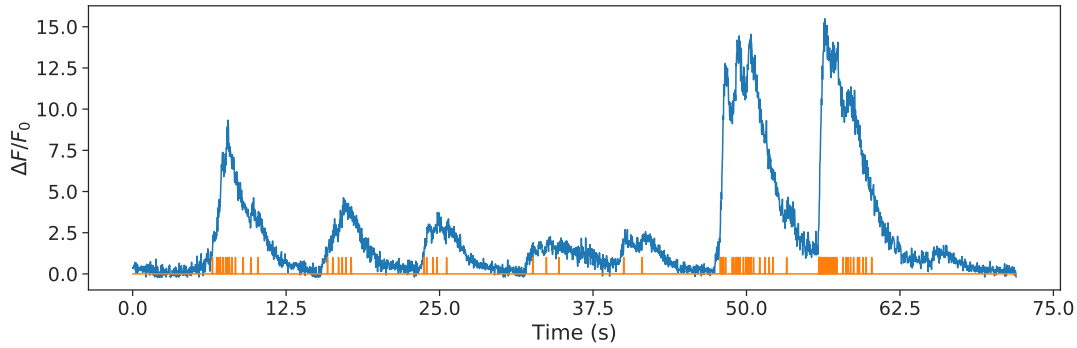


Figure 5.6: Example $\Delta F/F_0$ Ca^{2+} fluorescence trace and inferred spike train using the ML-Spike package

Neuronal activity in the form of binary spike trains can now be modelled using the statistical models presented earlier in this thesis. In addition, inferred spike trains can be used as an accurate readout of neuronal activity as they remove some of the variability that comes with analysing noisy fluorescence signals. This can be exemplified by the fact that stimulus decoding accuracy is higher when classifying stimuli using spike rates compared to raw fluorescence ($\Delta F/F_0$) signal, Fig. 5.7.

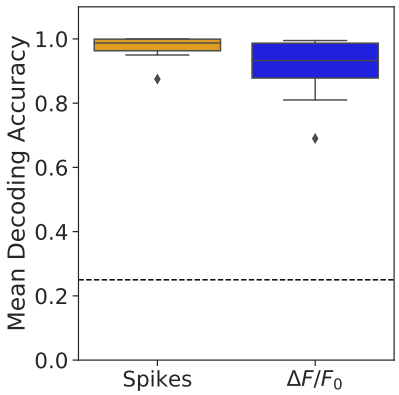


Figure 5.7: Stimuli decoding using inferred spikes and Ca^{2+} fluorescence

5.3 Results - Characterisation of Visual Encoding

5.3.1 Neuronal Population Activity

An initial step for the quantification of visual stimulus representation was to compute the average neuronal activity, in terms of mean $\Delta F/F_0$, exhibited within stimulation

periods and in darkness. The experimental group consisted of 14 animals ($N_{WT} = 8$, $N_{HET} = 6$), the number of neurons imaged in each animal field of view was $\mathbf{N}_{WT}^{FOV} = [124, 191, 213, 184, 183, 118, 170, 132]$ and $\mathbf{N}_{HET}^{FOV} = [200, 254, 169, 96, 175, 158]$. Fig. 5.8 shows the average fluorescence during phase reversal stimuli, intermittent grey screen, and darkness for each genotype (binocular stimulation groups only). On average, at the population level, there were no significant differences observed between the WT and SynGAP$^{+/-}$ animal groups.

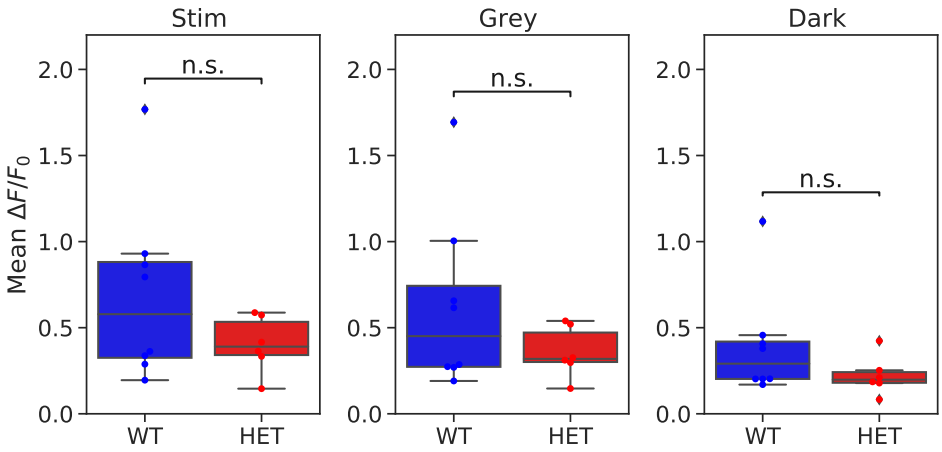


Figure 5.8: Mean $\Delta F/F_0$ per animal for each genotype during phase reversal, grey screen and darkness ($N_{WT} = 8$, $N_{HET} = 6$). n.s. indicates that groups were not significantly different ($p > 0.05$, t-test).

As individual neurons are tuned to different stimulus types, such as a particular direction or orientation, it was necessary to define a measure of visual responsiveness, in order to identify those neurons within the imaged population that were responding to visual stimulation. Many different methods for defining visual responsiveness exist [109, 110, 111], the one utilised here was developed by the Rochefort lab, and was selected in order to handle the high variability observed in response properties. The method involved bootstrapping the recorded trial responses during stimulus and grey-screen periods in order to obtain distributions for the mean responses during these two periods. The visually responsive cells were then taken to be those with significantly different stimulus and grey-screen response distributions ($p < 0.05$). Based on this criterion, the percentage of visually responsive cells was calculated per animal and compared across genotype, Fig. 5.9, with results indicating a high variance of values across animals that is not significantly different between animal groups.

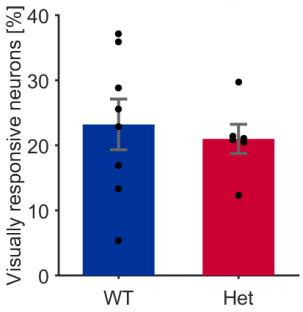


Figure 5.9: Percentage of visually responsive cells for each genotype

The population averages were again computed only this time considering those neurons deemed visually responsive, again revealing no significant differences between genotype, Fig. 5.10.

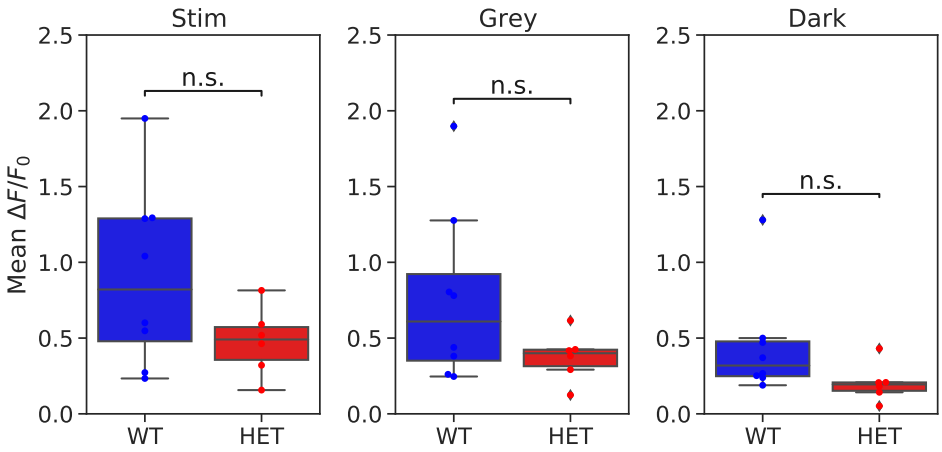
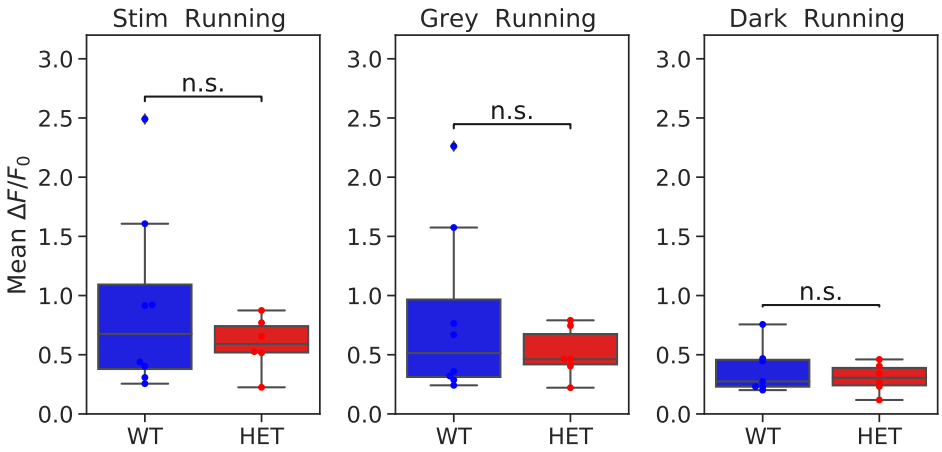
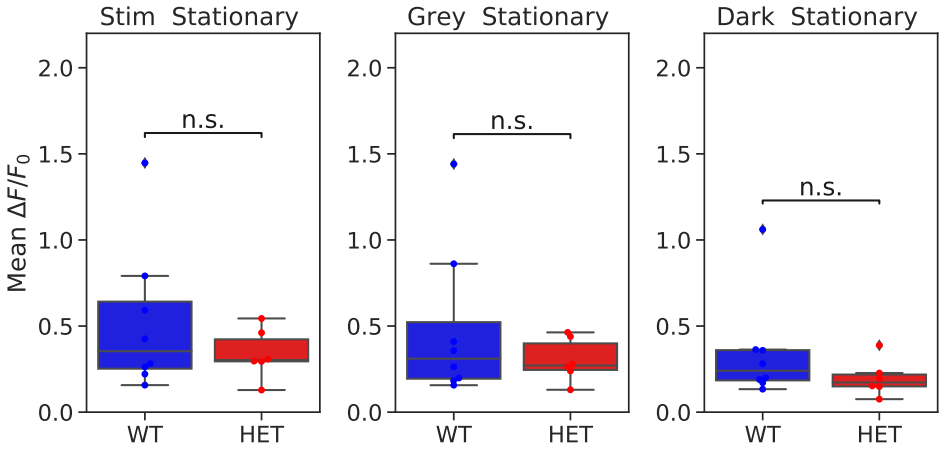


Figure 5.10: Mean $\Delta F/F_0$ per animal for each genotype during phase reversal, grey screen and darkness, only considering visually responsive cells

As the animals are free to move on a treadmill, another factor that affects neuronal activity is locomotion. When an animal is moving, its overall activity is boosted, with some cell activity coupled to this motion more than others (see Fig. 5.2(b) for example). It is then of interest to investigate whether this locomotion modulated activity is different in each genotype. Again, mean activity is computed for stim, grey and dark periods, only this time separating further the periods of locomotive and stationary activity, Fig. 5.11. Average $\Delta F/F_0$ was again comparable across genotype.



(a)

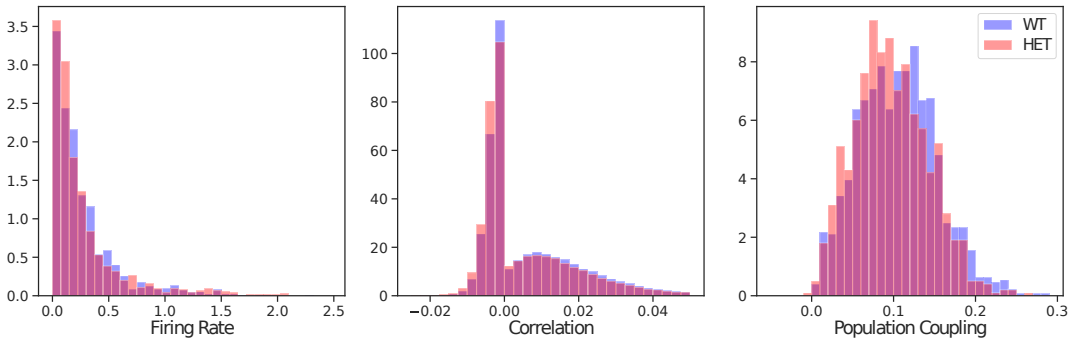


(b)

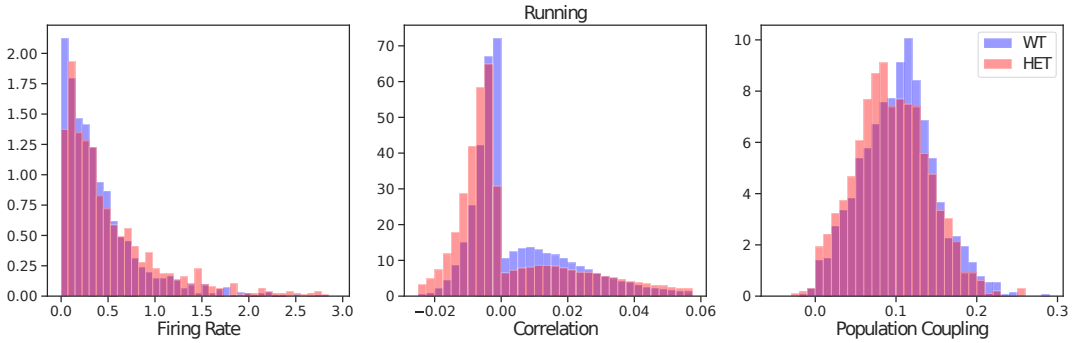
Figure 5.11: Mean $\Delta F/F_0$ per animal for each genotype during phase reversal, grey screen and darkness, during (a) locomotive and (b) stationary periods

5.3.2 Spiking Activity Measures

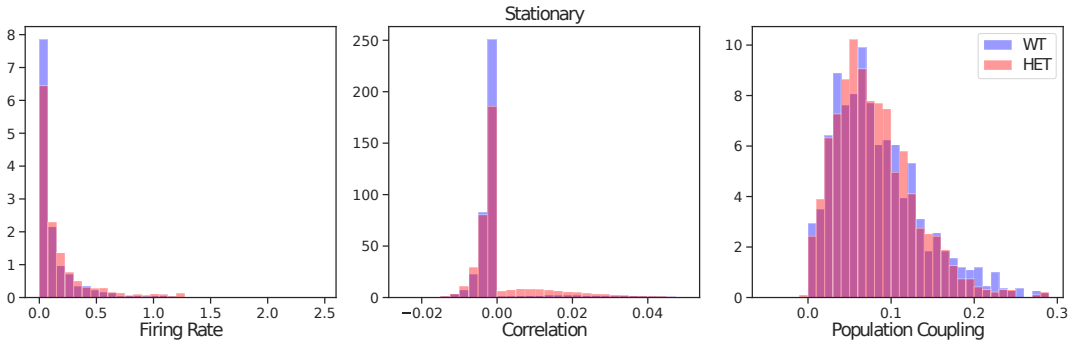
To support the network measures from mean Ca^{2+} fluorescence values the individual neural firing rates, correlations and population couplings were calculated using the inferred spiking activity. Fig. 5.12 shows the distributions of these values during stimulation grouped together over all animals per genotype.



(a)

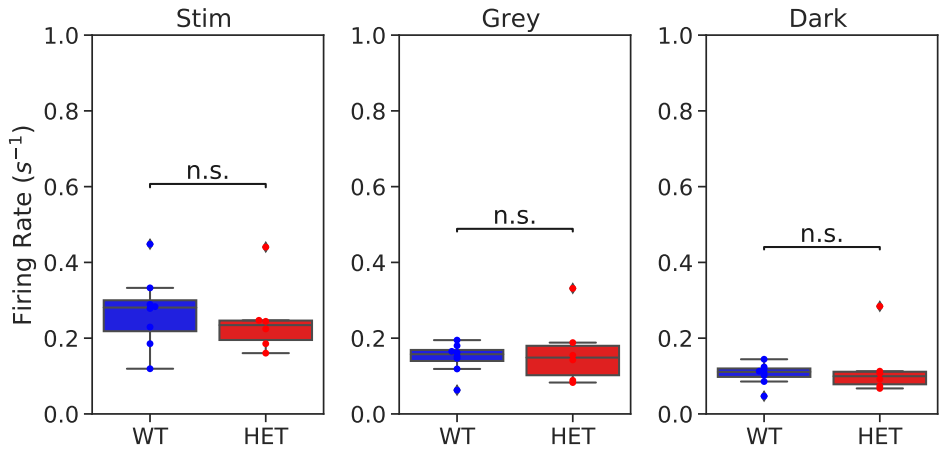


(b)

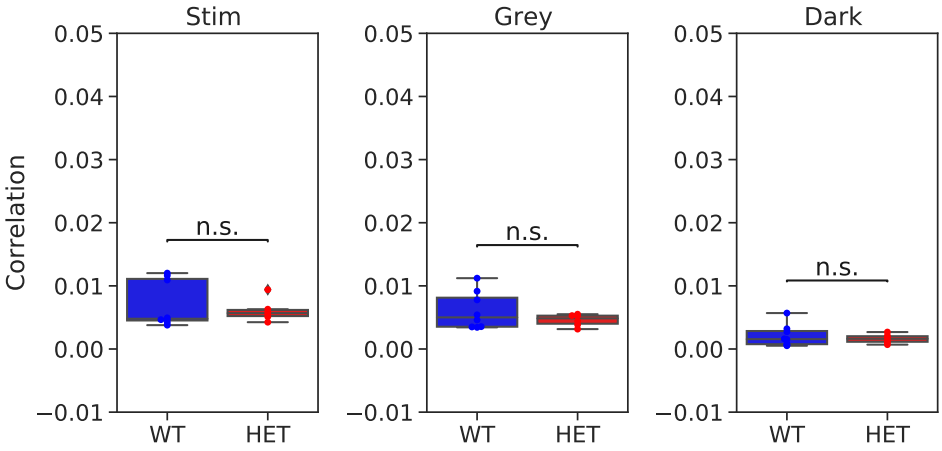


(c)

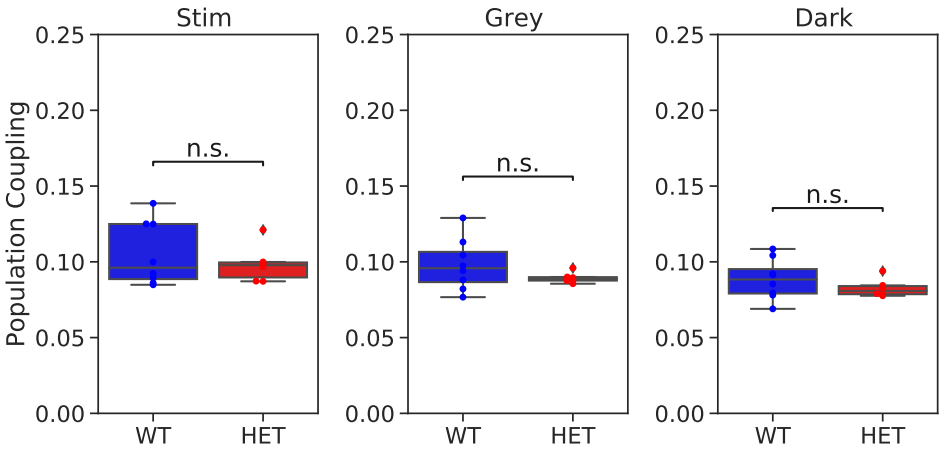
Figure 5.12: Distributions of firing rates (bin size = 0.075), pairwise correlations (bin size = 0.0025) and population couplings (bin size = 0.01) during stimulation for (a) all (b) locomotive and (c) stationary periods. Data pooled together over all animals in each genotype



(a)



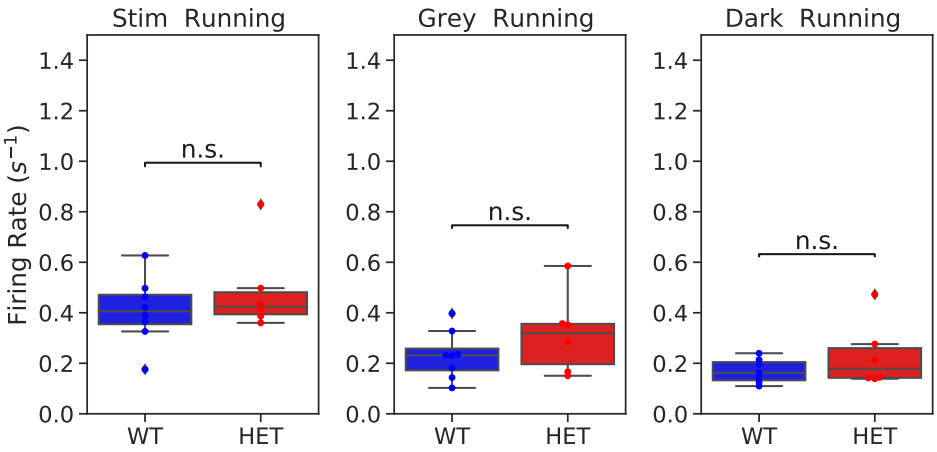
(b)



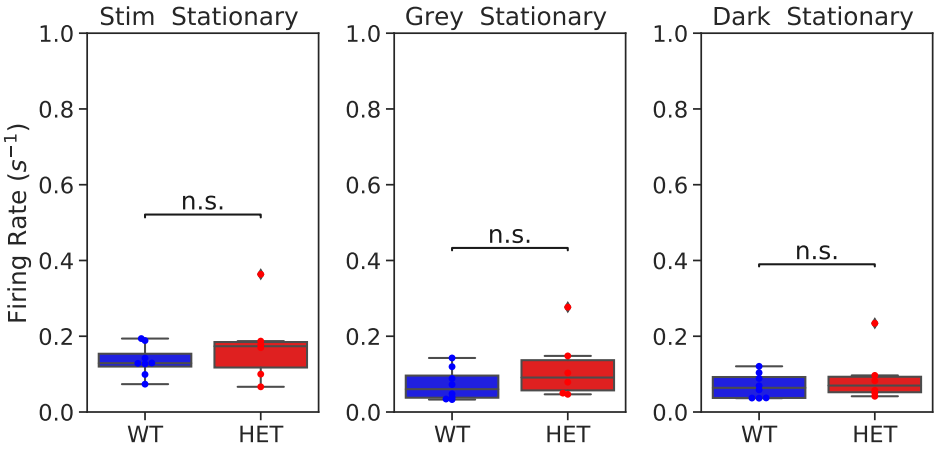
(c)

Figure 5.13: Mean (a) firing rate, (b) correlation coefficient and (c) population coupling per animal for each genotype during phase reversal, grey screen and darkness

Overall, distributions were similar, though a higher proportion of WT neurons exhibited higher correlation and population coupling during locomotive periods, compared with the HET group. Mean values for firing rate, correlation and population coupling were then computed per animal and compared across genotype, Fig. 5.13. Mean animal values again showed no significant difference between the genotypes. These measures were again computed on the separated locomotive and stationary signal, Fig. 5.14-5.16.



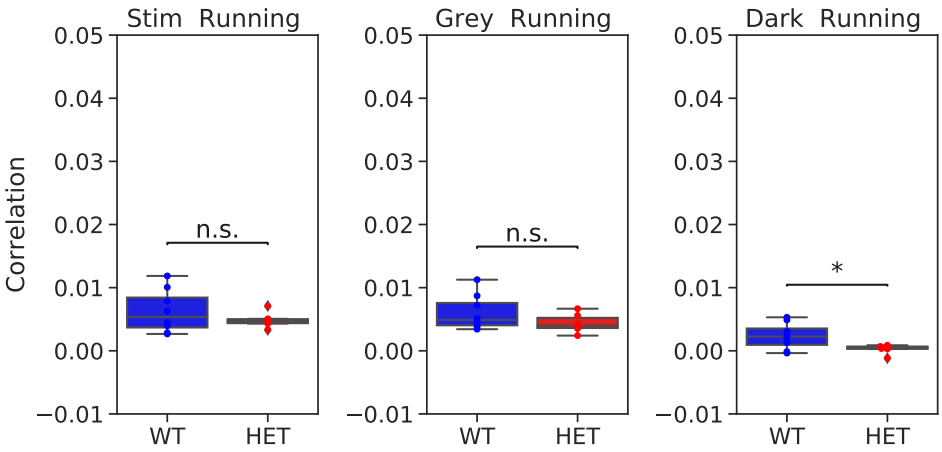
(a)



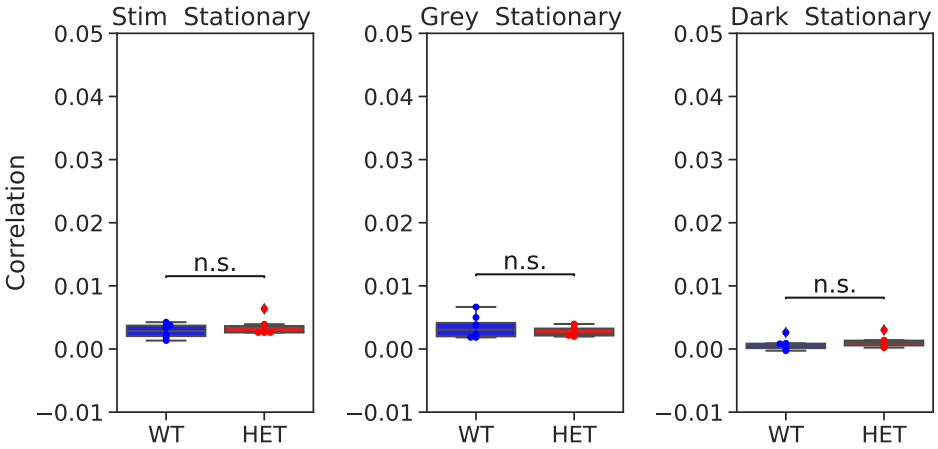
(b)

Figure 5.14: Mean firing rate per animal for each genotype during phase reversal, grey screen and darkness, during (a) locomotive and (b) stationary periods

The only significant difference was observed between correlation and population coupling during locomotive periods in darkness.

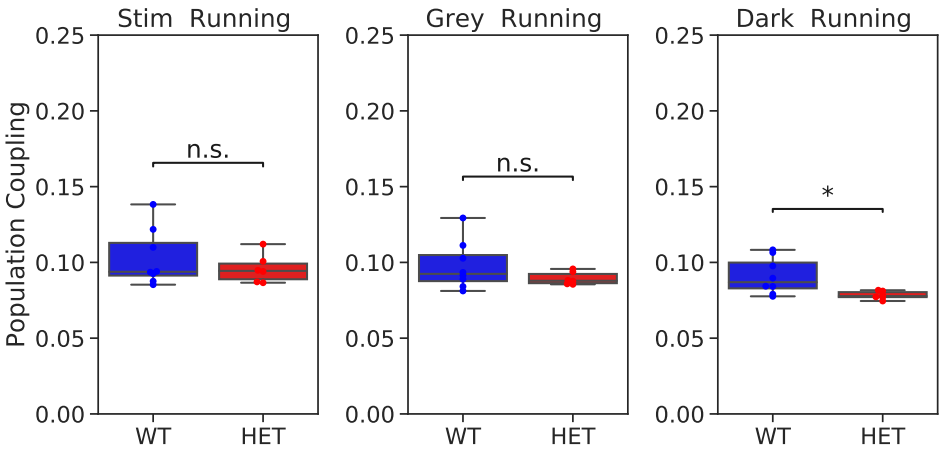


(a)

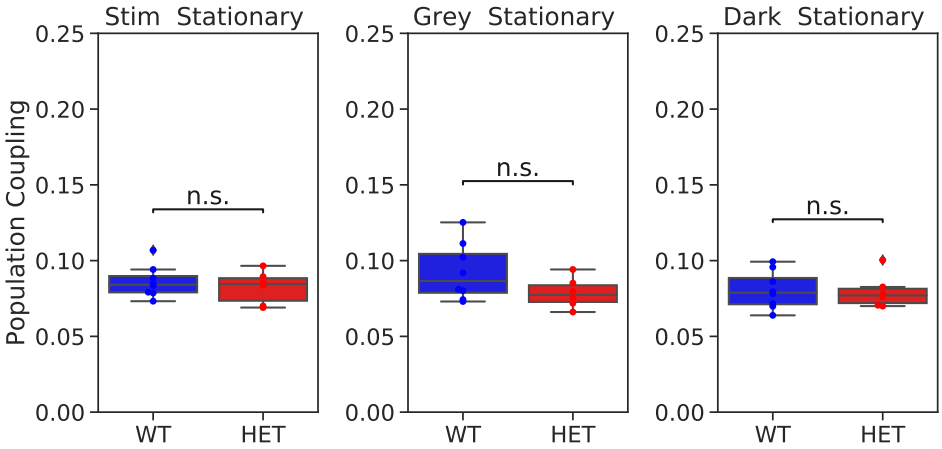


(b)

Figure 5.15: Mean correlation per animal for each genotype during phase reversal, grey screen and darkness, during (a) locomotive and (b) stationary periods. One star (*) indicates that groups were significantly different ($p < 0.05$, t-test).



(a)



(b)

Figure 5.16: Mean population coupling per animal for each genotype during phase reversal, grey screen and darkness, during (a) locomotive and (b) stationary periods

In conclusion, base measures of spiking activity were all comparable between the two animal groups, suggesting that differences in coding capabilities were not expressed at the level of population wide measures.

5.3.3 Orientation Selectivity

Many neurons in the visual cortex are selective to particular stimulus features, such as orientation and direction. In order to quantify this selectivity it is necessary to define a selectivity measure. As the stimuli used here is phase-reversing, the only feature that

varies is the orientation, so an orientation selectivity index (OSI) is defined using a neuron's mean response to each orientation.

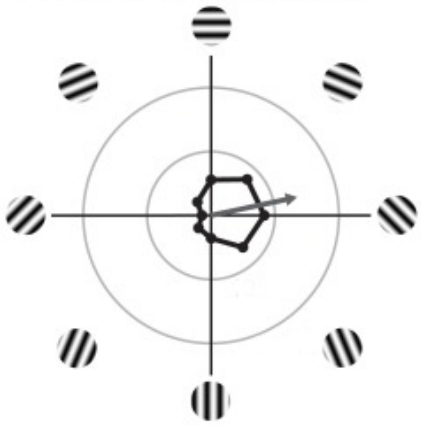


Figure 5.17: Mean response of a neuron to each stimulus orientation plotted in polar coordinates [112]

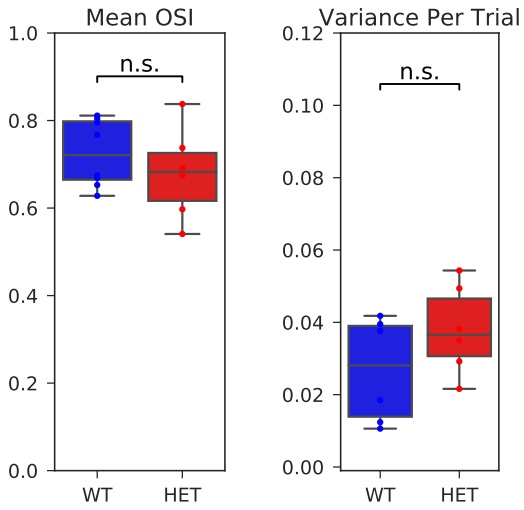
The OSI can be calculated by simply normalising a neuron's response to each orientation, however, a more sophisticated technique involves a quantity related to the circular variance of responses [112], and is calculated as follows:

$$OSI = \left| \frac{\sum_k R(\theta_k) \exp(2i\theta_k)}{\sum_k R(\theta_k)} \right| \tag{5.1}$$

where $R(\theta_k)$ is the mean response of a neuron to the k^{th} orientation. This method converts a cell's response to each orientation into polar coordinates and normalises. The value given by (5.1) represents the strength of selectivity for that neuron. Looking at Fig. 5.17, the preferred angle Θ_{Pref} (grey line) can also be calculated by summing the vectors given by the averaged responses.

Taking neural responses over concatenated trials in the binocular stimulation group yielded an overall OSI measure for each neuron. As locomotive periods varied significantly among animals, only stationary data was considered when computing OSI. This was to remove biases induced from locomotion during certain stimuli (e.g., if an animal happened to be running during a 45 degree presentation its activity would be boosted compared to other orientations and its OSI may not reflect the true selectivity). Additionally, non-responsive and weakly selective neurons (OSI ≤ 0.3) were not considered for analysis. Fig 5.18(a) shows mean OSI per animal grouped by genotype, demonstrating no significant difference between the average OSI measures. Another quantity

of interest was OSI reliability, estimated by calculating the variance of a neuron's OSI over individual trials (as opposed to concatenating trials), Fig. 5.18(b). While OSI variance was slightly higher on average in the SynGAP$^{+/-}$ animals, this difference was not statistically significant.



(a) (b)

Figure 5.18: (a) Mean OSI and (b) mean variance in OSI value over trials per animal, grouped by genotype. Only visually responsive cells and cells with an OSI > 0.3 were considered

Using the preferred angle (Θ_{pref}) obtained from the OSI calculation it is possible to calculate the matched preference between right and left eye ($|\Theta_{prefR} - \Theta_{prefL}|$). This is thought to be an indication of developmental plasticity in the brain, with the idea being that neurons should elicit responses to similar orientations when presented with stimuli in the contralateral and ipsilateral eyes [113]. Fig. 5.19 demonstrates this mean difference per animal, again with no significant difference observed between genotypes.

Results from the OSI calculations were somewhat surprising. Due to the known developmental and physiological differences between the two genotypes, there was an expectation for discrepancies in the encoding properties of the neuronal networks, whereas mean OSI, OSI reliability and matched preference were all comparable between WT and SynGAP$^{+/-}$ animals. One possible reason for such similarities may be the simplicity of the stimuli - phase reversal gratings of 4 differing orientations may not contain enough richness to illicit responses in which encoding deficits become apparent within the HETs.

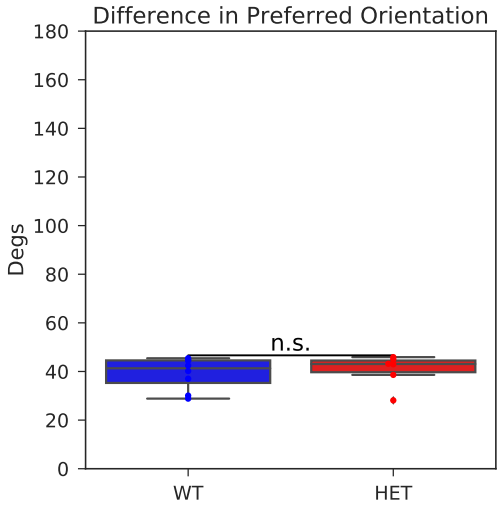


Figure 5.19: Mean difference in preferred orientation between right and left stimulation groups per animal for each genotype

5.3.4 Difference in Averaged Response to Stimuli

To focus in further on the characteristic features of the two groups of animals the average response to the stimulus was investigated. Here, each neuron's spiking activity during every 3 second stimulus presentation was binned and normalised to the maximum value, these signals were then averaged together to give a characteristic response for each genotype. Fig. 5.20 shows the averaged response of all neurons to all stimuli. Results indicate a stronger average response earlier from stimulus onset presentation in the HETs compared to a more prolonged response in the WTs.

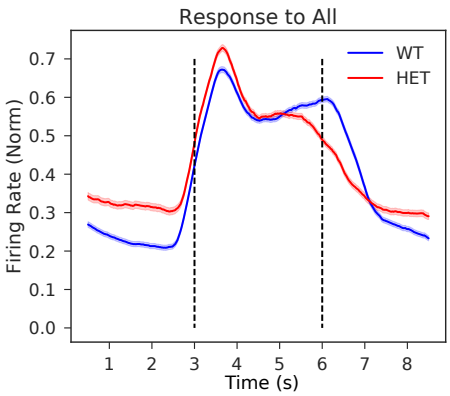
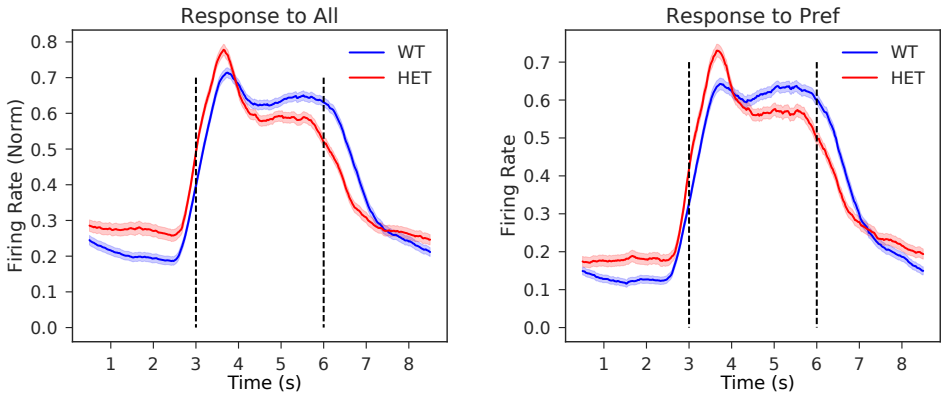


Figure 5.20: Mean response of each genotype to all stimuli for all neurons. Dashed lines indicate the onset and offset of the stimulus

Next the procedure was repeated using only visually responsive cells, plotting the average response to all stimuli as well as the preferred stimulus (the stimulus that evoked the highest response on average for a given neuron), Fig. 5.21. The resulting plots give a cleaner characterisation of neural responses and amplify the differences between the two genotypes. Results indicate a more varied distribution of response times in the WT animals profile - although a peak is also observed early in the presentation there are a large proportion of neurons that respond later when compared to the HETs.

As this is an average over all trials results could be an indication of habituation of certain neurons and so it is of interest to examine this average response across trials.



(a) (b)

Figure 5.21: Mean response of each genotype to (a) all stimuli and (b) the preferred stimulus only including visually responsive cells. Dashed lines indicate the onset and offset of the stimulus

Fig. 5.22 shows the average response of all neurons to all stimuli for the first 10 presentations during a recording session. As expected, there is a noticeable decrease in mean response after each presentation indicating a natural habituation to the stimulus. This decrease is most visible over the first 3-4 presentations, after which responses converge and stay around a baseline level. However, these response properties are similar between genotypes.

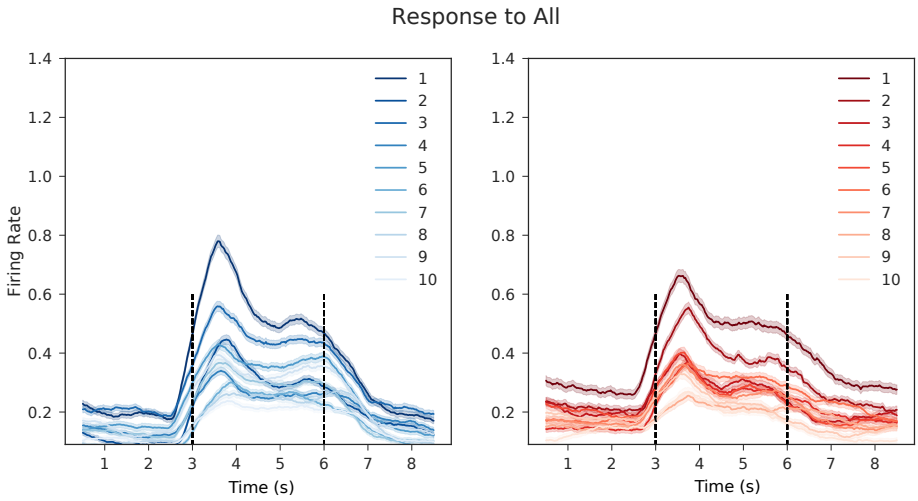


Figure 5.22: Mean response of (a) WT and (b) SynGAP$^{+/-}$ to all stimuli for all neurons at each presentation (shown in legend). Dashed lines indicate the onset and offset of the stimulus

Again this was repeated taking only visually responsive neurons into consideration and plotting mean neural responses to all stimulus as well as the preferred (Fig. 5.23-5.24). While the decrease in mean response with presentation is clear in both genotypes, there is a difference in the response shape. The WT animals demonstrate a clear peak early within the stimulation time for the initial presentation but as the presentations are repeated the average response begins to level, indicating a habituation in a proportion of neurons that were giving strong responses early in the window. The HET animals however seem more rigid in their responses, while they do show a clear habituation in magnitude across trials the overall shape of these responses remains constant - a clear peak early in the trial that levels after 1-2 seconds.

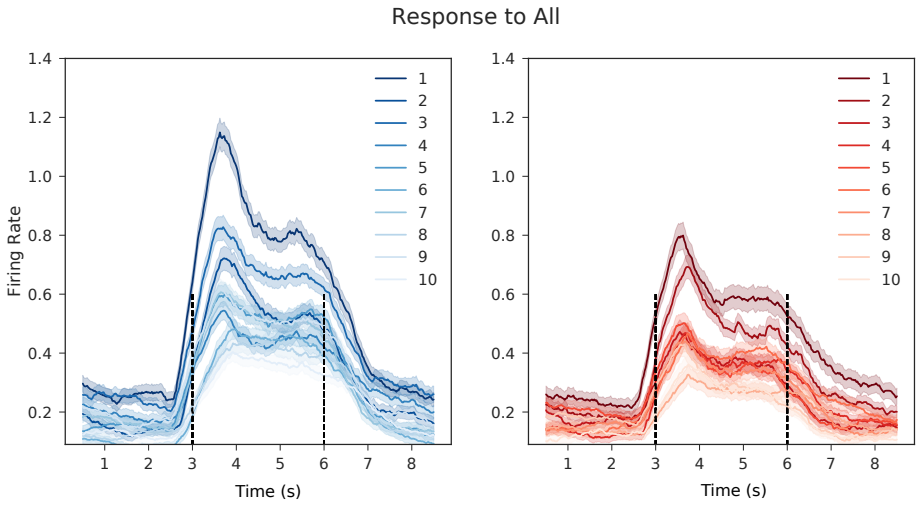


Figure 5.23: Mean response of (a) WT and (b) SynGAP$^{+/-}$ to all stimuli for visually responsive neurons at each presentation (shown in legend). Dashed lines indicate the onset and offset of the stimulus

This difference in shape can be quantified by calculating the difference between peak response and median response during the presentation window using the normalised traces seen in Fig. 5.25.

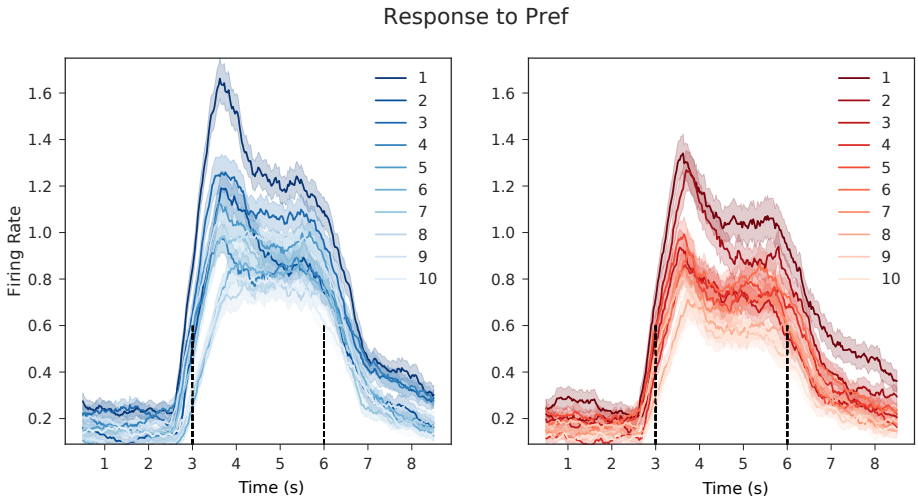


Figure 5.24: Mean response of (a) WT and (b) SynGAP$^{+/-}$ to preferred stimulus for visually responsive neurons at each presentation (shown in legend). Dashed lines indicate the onset and offset of the stimulus

Fig. 5.26 demonstrates this difference clearly, showing that the HET animals exhibit a higher difference between peak and base response during the stimulus presentation.

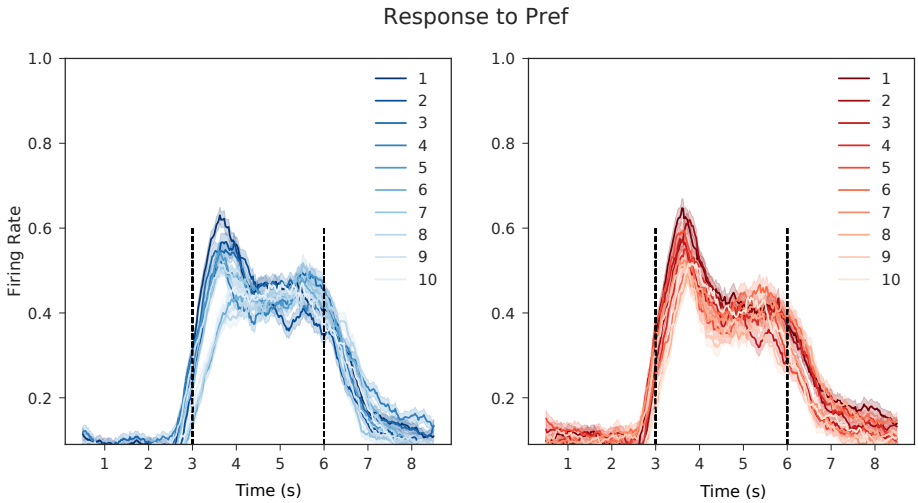


Figure 5.25: Mean response of (a) WT and (b) SynGAP$^{+/-}$ to preferred stimulus for visually responsive neurons normalised to the maximum response, at each presentation (shown in legend). Dashed lines indicate the onset and offset of the stimulus

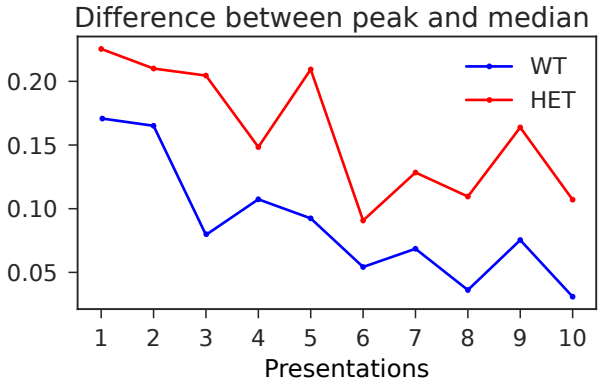


Figure 5.26: Difference between peak and median of normalised mean response to preferred stimulus

Overall, differences were present in the characteristic response properties of the two animal groups, with a stronger habituation effect visible in the WT animals. This difference, however, could be further explored and quantified with the use of a more sophisticated stimulation protocol.

5.3.5 Population Entropy

Aside from base measures of responses, the spiking activity of the animals was also modelled using the population tracking model (Sect. 2.5). As a reminder, the model

fits individual firing probabilities $p(x_i)$, the synchrony distribution $p(k)$, and the conditional probabilities that a neuron is active given there are K other active neurons ($p(x|k = K)$, in order to approximate a low dimensional measure of the firing pattern distribution (the population entropy). As discussed previously in Sects. 2.5.1 and 4.2.2, the population entropy offers a network wide measure of uniformity of responses, and can also be interpreted as an upper limit on the amount of information a population is able encode [70]. This latter definition is of particular interest when quantifying the ability of a network to represent visual stimuli through its responses and has been used in previous work to characterise coding capabilities in the mouse neocortex across development [33].

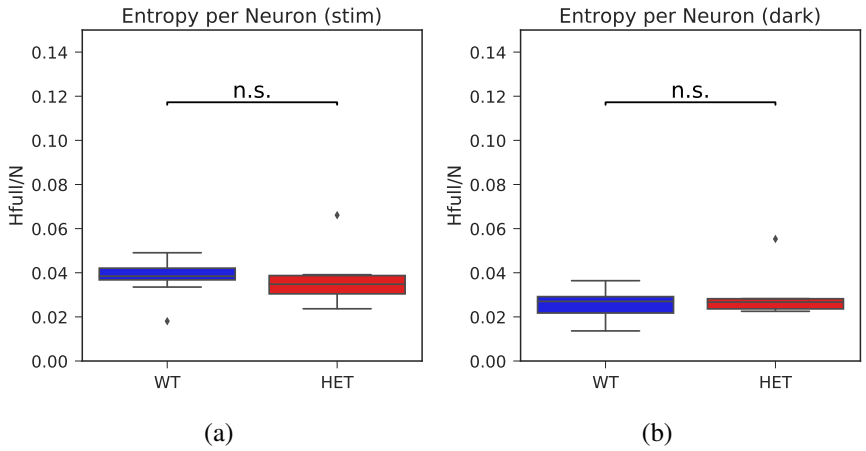


Figure 5.27: Entropy of the two genotypes during stimulation and in darkness

Fig. 5.27 shows the entropy per neuron for each of the two animal groups for both evoked and spontaneous activity, evidently the two quantities are on average not significantly different, supporting previous findings that suggest neuronal population function in V1 is comparable between the two genotypes. Another simple model of the population is the independent model, which assumes that the model distribution arises solely from individual neural firing rates (all neurons are independent). Based on this assumption, model entropy can be easily computed, and Fig. 5.28 demonstrates the correlation between this independent model entropy and the entropy calculated from the population tracking model. The strong correlation indicates that contributions of population coupling and correlation do not influence the model entropy significantly and that due to low correlations in the typically sparse activity, the independent model is a satisfactory fit to the data.

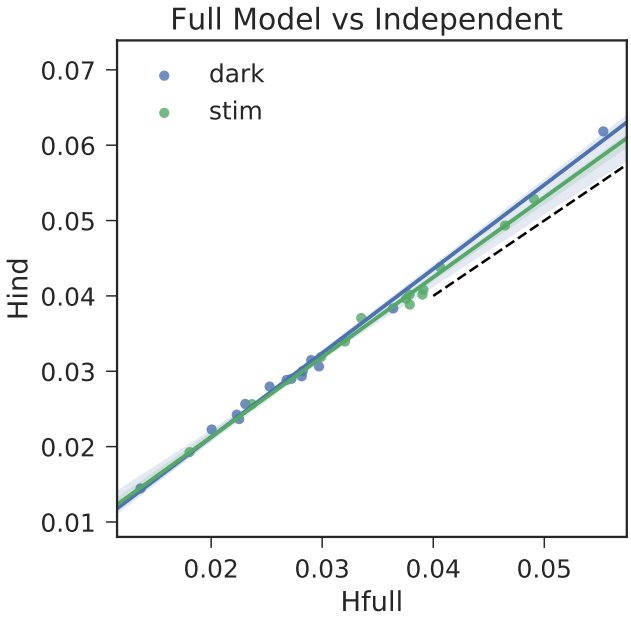


Figure 5.28: Entropy given by full model correlates strongly with that of the independent model, indicating that correlation and population coupling contribute weakly in the model

5.4 Results - Cortical Plasticity Induced by Monocular Deprivation

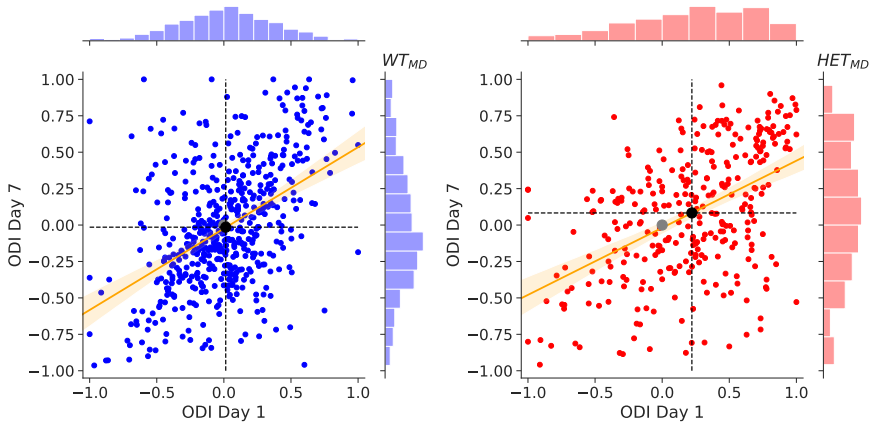
Having shown the similarities in visual encoding between the two genotypes at the population level, it was then of interest to investigate whether the known physiological deficiencies in the SynGAP$^{+/-}$ animals could be observed via another circuit wide mechanism, plasticity. The following section analyses neural responses before and after a 7-day induced plasticity paradigm (MD), using modelling techniques presented in the earlier chapters of this thesis.

5.4.1 Mean Response and Ocular Dominance

Ocular dominance is a measure of eye preference and is quantified using the ocular dominance index (ODI), which is defined as follows

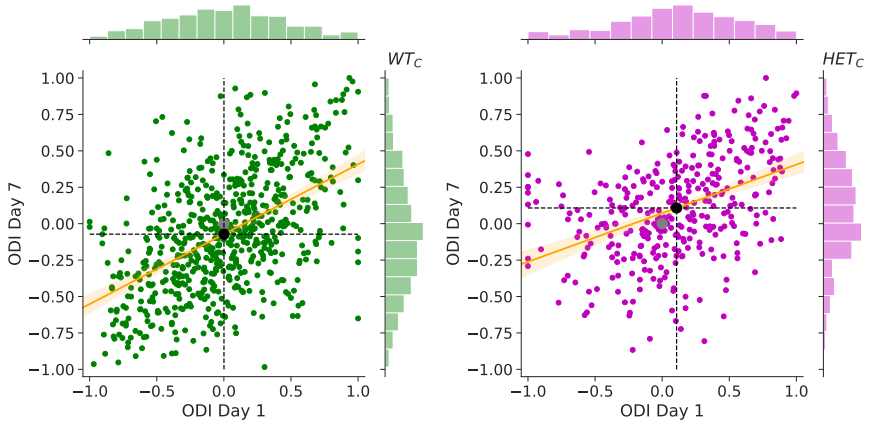
$$ODI = \frac{R_{Right} - R_{Left}}{R_{Right} + R_{Left}} \tag{5.2}$$

where R_{Right}, R_{Left} are the mean responses of neurons during the right and left monocular stimulation groups. Fig. 5.29 shows the ODI computed for each neuron within the two genotypes before and after MD.



(a) (b)

Figure 5.29: Ocular dominance index for WT and HET MD animal groups before and after 7 day MD period. Mean shift -0.0285 and -0.1383 respectively

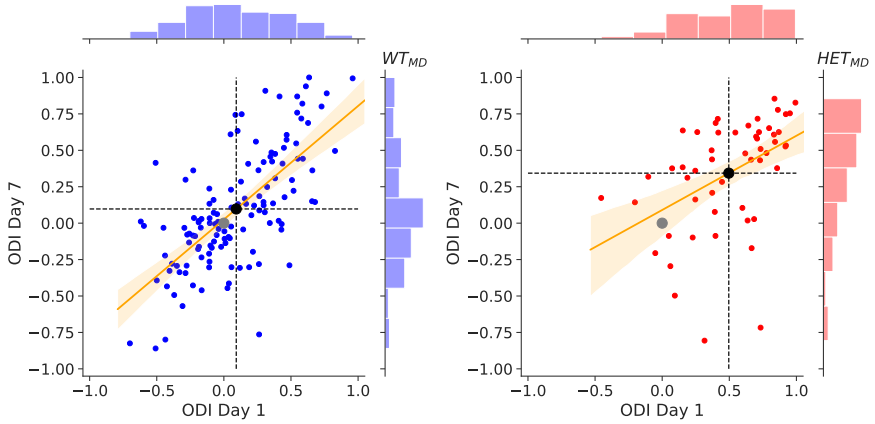


(a) (b)

Figure 5.30: Ocular dominance index for WT and HET control animal groups before and after 7 day control period. Mean shift -0.0728 and 0.00029 respectively

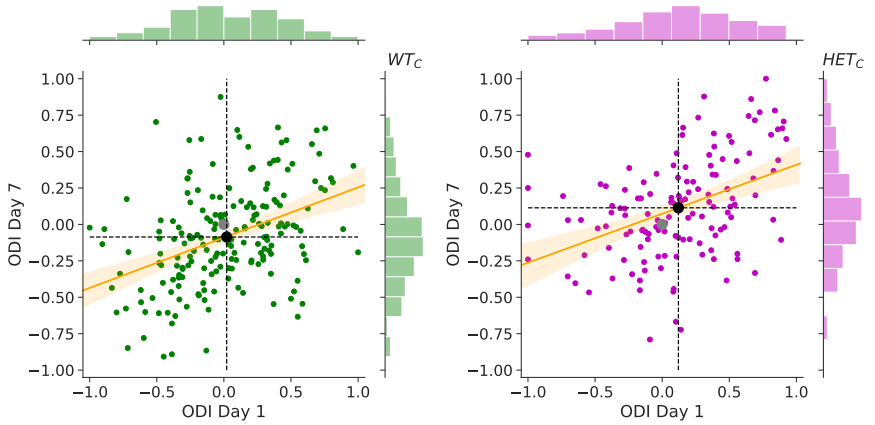
Evidently, the ODI measures are highly noisy across the 7 day MD/control periods, with only the HET MD group appearing to have successfully undergone a shift. This is likely due to the fact the experiment was undertaken on awake-behaving animals in

which many factors such as locomotion and attention can influence activity. Corresponding ODI shifts for visually responsive neurons are also shown in Figs. 5.31-5.32.



(a) (b)

Figure 5.31: Ocular dominance index for WT and HET MD animal groups before and after 7 day MD period, visually responsive cells only. Mean shift 0.00524 and 0.343 respectively



(a) (b)

Figure 5.32: Ocular dominance index for WT and HET control animal groups before and after 7 day control period, visually responsive cells only. Mean shift -0.1073 and -0.0068 respectively

Overall, a clear MD shift was not visible for the WT animals indicating that the experiment had failed to capture an induced plasticity effect outside the variability of activity due to other factors, suggesting that experimental design should be improved upon in future work (see discussion for more details).

5.4.2 Model Entropy and Divergence

The spiking activity before and after MD was then modelled in order to extract information on population entropy and compute divergence measures. This was done by both subsampling smaller populations to obtain Ising model fits, and modelling the full imaged ensembles using the population tracking model. In the case of inferred spiking activity, no binning procedure was required as the frame rate of image acquisition was $40Hz$ ($\delta\tau = 25ms$).

5.4.2.1 Ising Model

Similar to the procedure presented in Sect. 3.4.1, 20 groups of 10 neurons were sampled from each imaged population in order to obtain smaller subnetworks for Ising model fitting. The model entropy (per neuron) was computed for each group before and after MD, Fig. 5.33.

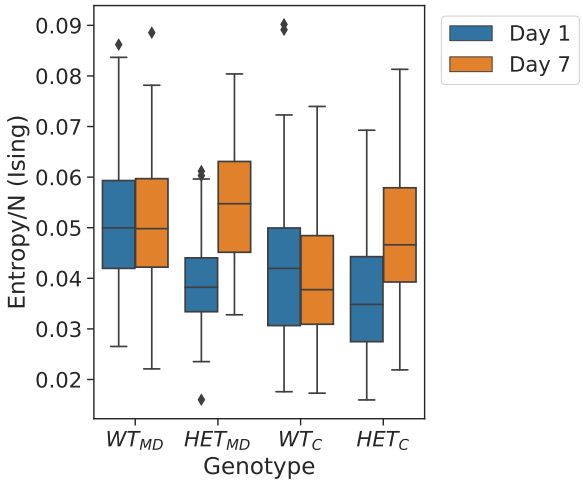


Figure 5.33: Entropy per neuron computed for model fits before and after 7-day period

Results indicate a conserved level entropy across the 7-day period for both WT groups, where as a significant jump in entropy in the HETs. This relation reflects the average rate observed for each animal, computed over the full imaged populations, Fig. 5.34

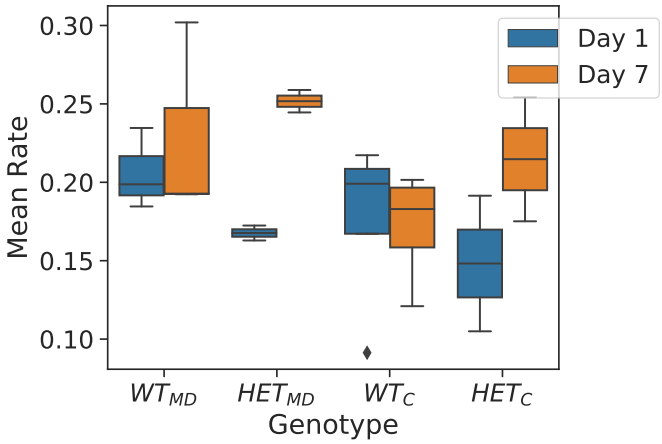


Figure 5.34: Mean rates computed per animal before and after 7-day period

Using the model fits, divergence measures were also computed between groups before and after the 7-day period, however, no significant differences were observed between genotypes or within MD and control groups.

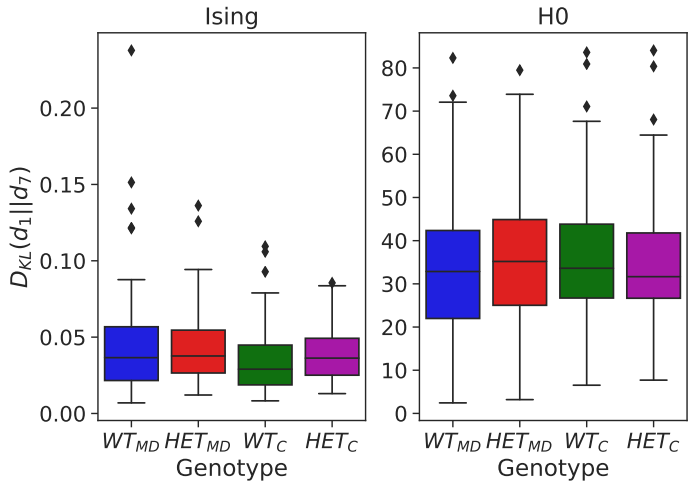


Figure 5.35: KL divergence computed between day 1 and 7 for the Ising model (right) and H0 model (left)

Models were also fit to activity from the left and right eye stimulation groups, to again compute a divergence between the two, before and after MD.

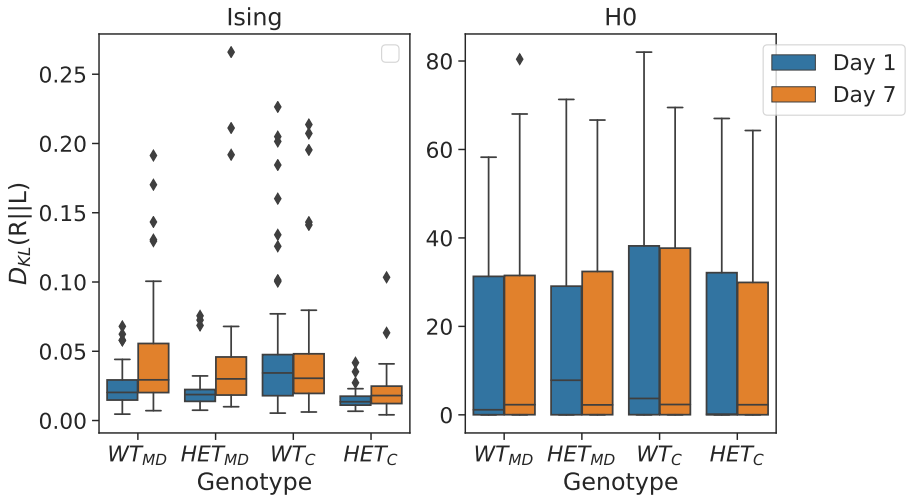


Figure 5.36: KL divergence computed between the right and left eye on day 1 and day 7, for the Ising, and H0 models

Although divergences between the right and left eye increased on day 7 for the MD groups and remained constant for the control groups, significant differences between genotypes were not visible.

5.4.2.2 Population Tracking Model

The analysis was then repeated using the population tracking model, taking the full neural populations from each imaged field of view as the training data. Entropy calculations mirrored those of the Ising model, again following the mean rates of each population, Fig. 5.37.

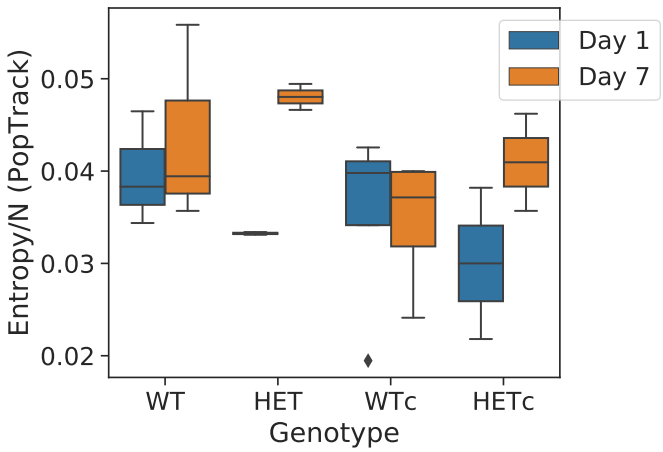


Figure 5.37: Entropy per neuron computed for model fits before and after 7-day period

Divergence measures indicate a larger difference over days for the MD groups compared with the control, although again these quantities were similar for WT and SynGAP$^{+/-}$ animals.

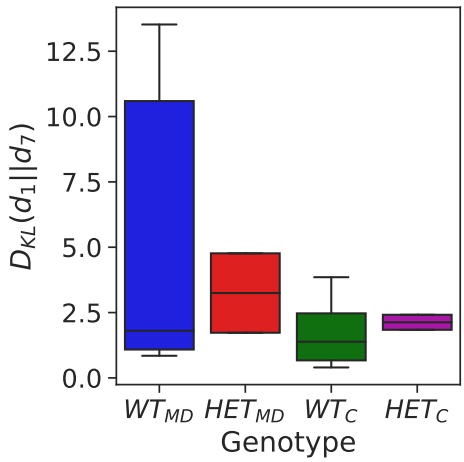


Figure 5.38: KL divergence computed between day 1 and 7 for the population tracking model

5.4.3 Diagonal Fisher Information Matrix

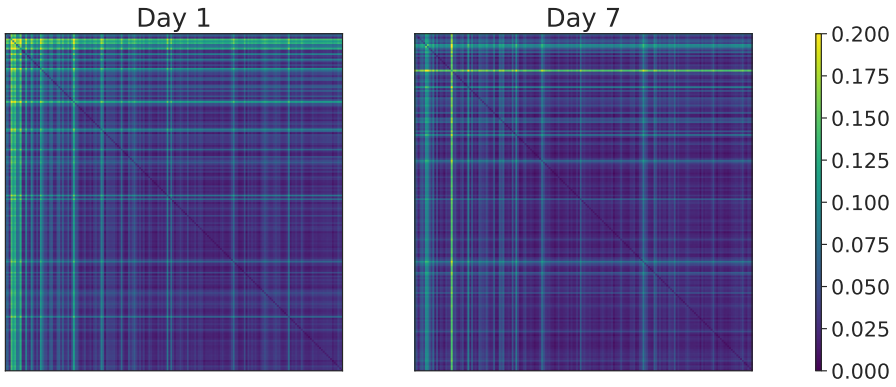
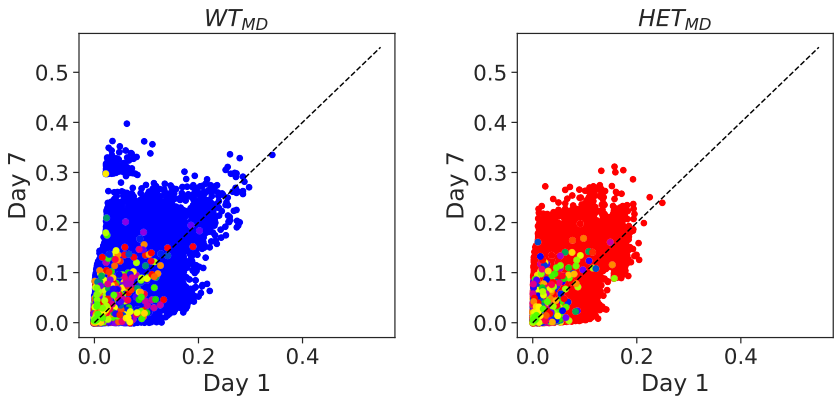


Figure 5.39: Reshaped FIM diagonal for an example recording on day 1 and 7 (WT animal, binocular stimulation)

As demonstrated in Fig. 5.39, parameter sensitivity is less conserved over days when compared to the cultured experiments. This is an expected result due to the highly variable nature of activity from awake behaving experiments, with many more external factors, in addition to more developed networks, combining to make a mode varied set of responses.

Figure 5.40: Reshaped FIM diagonal for the same animal on day 1, right vs left eye stimulation

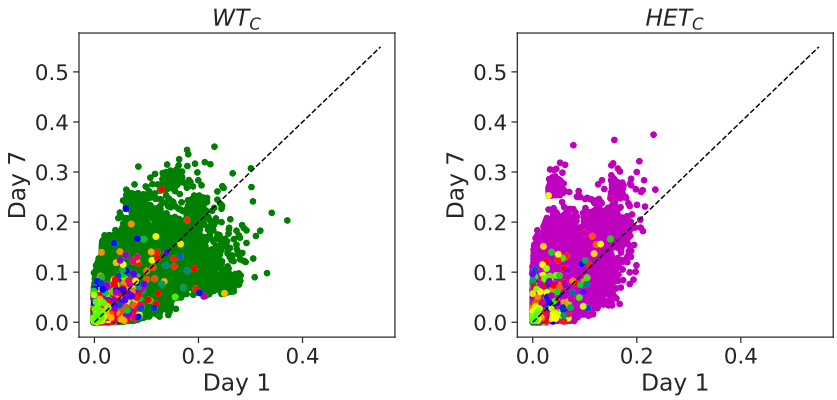
Similarly, diagonal FIMs show differences in structure between right and left eye stimulation groups, Fig. 5.40.



(a) (b)

Figure 5.41: Change in Fisher diagonal entries across days for the WT and HET MD animal groups. Entries corresponding to field parameters highlighted in colour

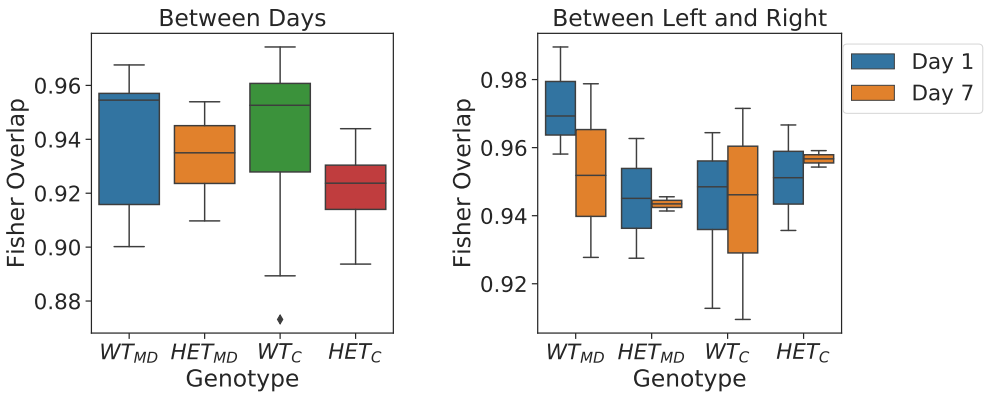
This difference in structure can be visualised for all stimulation groups across the 7-day period in Fig. 5.41 - 5.42, which clearly indicates the variable nature of neural activity across the experiment. Here, stability is vastly different to that of the neuronal cultures explored in chapter 4, with sensitivity fluctuating significantly within the populations.



(a) (b)

Figure 5.42: Change in Fisher diagonal entries across days for the WT and HET control animal groups (each coloured dot represents a neuron)

The scatter plots also failed to reveal any distinct characteristics pertaining to individual genotypes, which were further analysed using the Fisher overlap, Fig. 5.43. Results indicate a higher (though not significant) mean overlap between the WT animals across both the MD and control periods and also demonstrate a loss of similarity between the right and left stimulation activity for the WT MD group. Overall, network activity of the WT and SynGAP$^{+/-}$ animal groups were found not to be significantly different in response to the induced plasticity mechanism.



(a) (b)

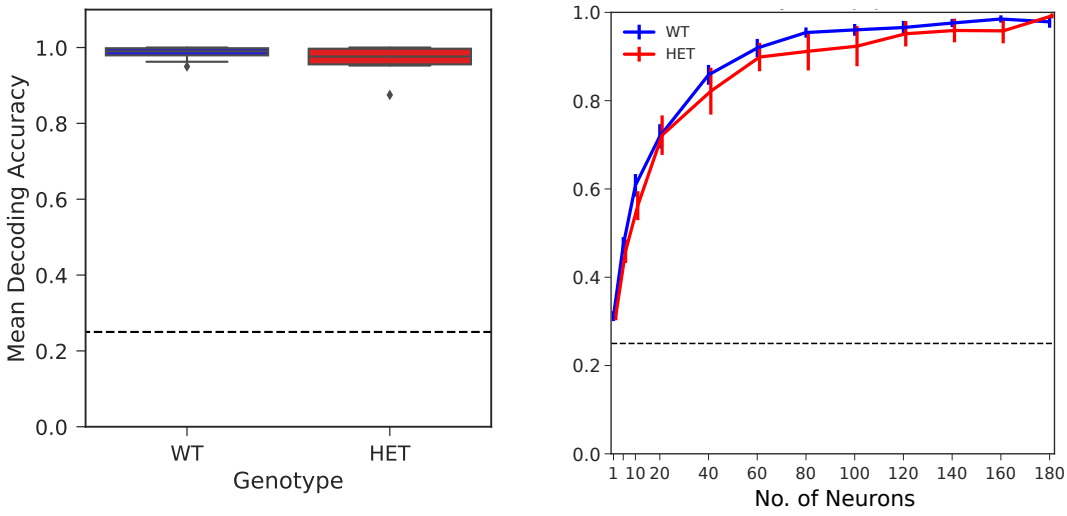
Figure 5.43: Fisher overlap of each experimental groups (a) between 7 day MD/Ctrl period and (b) between left and right eye stimulation groups

5.5 Discussion

Due to the recent findings on another mouse model of ASD, FMR1$^{-/y}$ [114, 115, 33] (linked to fragile X syndrome), and the convergence of the two genes (FMR1$^{-/y}$ and Syngap$^{+/-}$) on common hippocampal pathophysiology [116], there was an expectation that similar differences would be found between WT and SynGAP$^{+/-}$ mice. However, in response to the visual stimulation protocols utilised in this experiment, WT and SynGAP$^{+/-}$ animals did not show any significant differences in $\Delta F/F_0$ or spiking activity at the population level. This was demonstrated by the similar response properties exhibited within each genotype during different stimulation protocols (stim, grey and dark) as well as during stationary and locomotive periods. Possible reasons for such similarities may be the simplicity of the stimuli - phase reversal gratings of 4 differing orientations may not contain enough richness to elicit responses in which encoding deficits become apparent within the HETs. Measures of orientation selectivity were also comparable across the genotypes, both in terms of the average strength of the neuron's preference, and the variability of this value across experimental trials. This similarity in encoding capability was further demonstrated when comparing the classification accuracy of trained decoders using spiking activity from each animal group, Fig. 5.44. Evidently, accuracy levels were comparable ($> 90\%$), indicating that information on stimulus orientation was similarly prevalent in the neural activity of both genotypes.

Some slight differences did emerge however, when comparing the characteristic response of each genotype during stimulus presentation (to all stimuli or the preferred). This difference suggested a lack of habituation in the SynGAP$^{+/-}$ responses, demonstrated by a rigidity in response shape throughout all stimulus presentations.

Overall, apart from some minor observed discrepancies, major differences in visual stimulation encoding within the primary visual cortex were not found between the two genotypes. As mentioned, this could in part be due to the simplicity of the stimulation protocol, however, it also could be possible that the known physiological impairments that are developed in the SynGAP$^{+/-}$ animals are manifested at the network level in higher order areas of the visual cortex.



(a) (b)

Figure 5.44: (a) Mean decoding accuracy for (a) full populations and (b) subpopulations of increasing N

Additionally, plasticity mechanisms within the visual cortex were investigated using monocular deprivation. However, differences between the two genotypes in response to this experimental paradigm were again not observed, both in terms of spiking response measures and model analysis. Diagonal FIM measures revealed the fluctuating nature of network activity across the experiment, reflecting the highly variable behaviour of activity from awake behaving experiments, in which many more external factors such as attention and locomotion combine to create a far more varied set of responses. Furthermore, the slow habituation of responses to stimuli over trials creates an additional interference to measuring ODIs, especially as this habituation is variable among animals. This is in line with results from Bonhoeffer et al. [105], who report a high variability in ODI values for awake behaving animal experiments, suggesting that MD may be better measured under anesthesia. Additionally, analysis carried out by the Rochefort lab on the same animal groups using $\Delta F/F_0$ signal also revealed no difference in plasticity between the two genotypes [117]. These factors, coupled with the relatively low number of animals per genotype (divided across MD and control groups), suggest that a higher number of animals are necessary in order to offer evidence for a definite hypothesis on the plasticity mechanisms in SynGAP$^{+/-}$ mice, and remains a potential avenue for future work.

Chapter 6

Discussion

The research presented in this thesis has sought to further the understanding of systems neuroscience both by expanding the assortment of modelling techniques used for the large-scale analysis and interpretation of neuronal network activity, and by applying these tools to novel experimental settings, in order to characterise circuit reorginisation in cultured networks and classify visual coding impairments in a mouse model of ASD. The subsequent sections will provide a concise summary, in addition to detailing some of the limiting factors that were involved and potential future avenues in which to continue this body of work.

6.1 Limitations

In chapter 3, a novel analysis of the Ising model was presented involving the renormalisation of pattern probability distributions. While results using simulated DG networks offered a successful proof of concept, a more extensive collection of recorded datasets is needed in order to fully explore the potential of the H0 model in discovering connectivity changes in neuronal networks. As stated in Sect. 2.6.2, neuronal cultures, in conjunction with MEAs, offer a controlled environment in which to monitor the large scale activity of *in vitro* neural populations. However, due to the nature of these networks, the activity exhibited by them does not fully capture the complexity of mature living networks found in the brain [48, 49], imposing a further limitation on the scope of the model's potential. In a similar manner, chapter 4 introduced the use of

a diagonal Fisher information analysis, which showed promising results in detailing large-scale network structure but was again constrained by the same cultured datasets. Both forms of analysis were again used to quantify plasticity effects in awake behaving mice using inferred spiking activity, however, due to the high variability of the data and other limitations in the experimental design (discussed shortly), the methods failed to offer any further insight into circuit mechanisms.

While the Ising model has been shown to accurately reproduce the entire pattern distributions of spiking networks by only fitting first and second order statistics, they do come with some disadvantages. These limitations, such as the stationary nature of the fit and size constraints are still present in the renormalisation. The Ising model has, as a result, fallen out of favour and has been surpassed by other models, such as the restricted Boltzmann machine (RBM), capable of fitting larger networks and higher-order correlations [38, 39]. RBMs fit the pattern distribution through the introduction of a hidden layer, giving the network conditional independence which allows for the efficient computation of visible unit probabilities for much larger networks. This, however, does come at the cost of a loss of intuition about the network structure as it is unclear as to what attributes of the data the hidden units are modelling, e.g. time dependence or higher order correlations, and so an analogous renormalisation here is not possible [40]. Like the Ising model, an RBM also only captures stationary spiking distributions, which is a significant setback due to the clear importance of temporal dynamics in spiking activity [118, 119, 120], and while efforts have been made to include temporal correlations into both these model via structural changes or otherwise (temporal-RBM [121, 122], kinetic Ising [29, 123, 124]), the additional computational cost incurred means that these models are not feasible for use on reasonably sized spiking networks. In order to capture these temporal dynamics another class of models is required, for example, generalized linear models (GLMs). GLMs are widely used to fit spiking neural data and are able to successfully capture temporal dynamics within spike trains through the use of linear-nonlinear filters with probabilistic spiking mechanisms [125, 126], making them an excellent choice for decoding tasks [127, 128, 129]. However, while these models have the advantage of modelling time dependencies, they are also difficult to interpret and do not scale well to larger networks (due to the fact they only model pairs). Another set of models that have enjoyed recent success in capturing these dynamics are recurrent neural networks (RNNs), which have been shown to have uncovered complex temporal dependencies within large neural populations [130, 131]. However, again, interpretation of such networks remains a significant issue due to their

'black box' methods of operation.

While the population tracking model manages to overcome the size constraints imposed by the combinatorial explosion of possible patterns, it also only captures a stationary distribution over them. This fact is exemplified in the original paper [32], in which spike trains sampled directly from the model (for a sufficient number of samples) failed to reproduce the temporal correlations found in the data. Additionally, these sampled spike trains only partially replicated the correlation structure of the original data, displaying only coarse features of this structure, compared to an almost exact match that can be obtained when using the Ising model (albeit for much smaller networks). This is an expected downside to using such an approximation, and even despite this fact, the population tracking model offers the most efficient estimation of entropy for such large networks while still maintaining a satisfactory level of accuracy. The diagonal Fisher approximation for larger networks maintains the assumptions of a pairwise maximum entropy model in an effort to overcome the size constraints imposed on them. It should be noted, however, that while this approximation can still be informative in some cases [44], the use of only the diagonal entries does come with a significant loss of information, detailing only the FIM intersection with the parameter axes (see Fig. 4.12), and so should be interpreted with caution. When comparing the diagonal of the FIM to the principal eigenvectors of the full FIM for smaller networks (Sect. 4.3), significant similarities were observed in the structure between the two, with the diagonal acting as a sort of 'silhouette' of the eigenvector (Fig. 4.13), preserving the shape but not the magnitude of the entries. Due to the fact that the FIM diagonals computed for larger networks maintain a similar sparse structure, it could be that corresponding network mechanisms characterised by these smaller network eigenparameters are operating on a larger scale. However, due to the number of approximations involved, it is not possible to say with confidence that this is the case, and would require further investigation.

Limitations affecting the Ca^{2+} imaging experiments arose largely from the difficulties of awake-behaving animal experiments. Across the experiments, data acquired from a percentage of the imaged animals was discarded due to quality concerns, such as blurred fields of view and drifting ROIs between stimulation groups, and so more data is again needed to strengthen findings. Additionally, a more diverse/complex visual stimulation protocol would be vital in exploring potential difference in visual stimulation representation in primary visual cortex. The current phase reversal gratings offer only 4 distinct stimuli and often elicit poor responses after a small number

of trials. This may have to do with the animal's attention to an uninteresting stimulus type, which could simply be extended to more complex alternatives (e.g. direction dependent drifting gratings), in order to reveal potential encoding deficiencies. Another limitation arises from the fact that certain factors that are highly influential on neuronal activity, for example stress or attentiveness, are not accounted for in the analysis and can bias results. As discussed in Sect. 5.5, this can especially influence monocular deprivation experiments, and has also been found in other large scale studies of MD in the field [105], suggesting MD may be more reliable when carried out on anaesthetised animals.

6.2 Future Work

Firstly, in relation to the cultured experiments, more datasets from different scenarios in which to test the presented methodologies would be an ideal extension to the application of the renormalised Ising model and diagonal Fisher approximation. In particular, controlled experiments specifically designed to include key manipulations of correlation structure could offer a better testbed for model exploration than experiments previously carried out with other purposes in mind. The addition of new datasets should also include a more diverse collection of activity, i.e. from *in vivo* experiments, to avoid the limitations that arise from cultured neuronal networks. More specifically to the Fisher approximation, due to the similarity in structure between the diagonal FIMs and the principal eigenvectors of the full FIMs of Ising models, an exciting next step would be to explore the parallels between sloppy model behaviour characterised in smaller networks [41] and the similar structure observed in the larger diagonal approximations.

For the SynGAP$^{+/-}$ experiment, more animals would be vital to add statistical power to the monocular deprivation findings. Additionally, a more sophisticated experimental paradigm would also be a simple and effective next step to take, for example, the straightforward introduction of direction into the stimulus would increase the amount of presented stimulus and include the distinction of direction sensitive cells from the orientation selective ones. Furthermore, Ca^{2+} imaging can additionally be used to image inhibitory interneurons in parallel with excitatory cells, which would add an interesting extension to the presented analysis and potentially reveal a difference in coding capabilities between the two genotypes, especially due to the fact that

perceptual impairments have been hypothesised to arise from an imbalance in excitation and inhibition in the cortex [33, 132, 133]. Currently, work within the Rochefort lab is being undertaken to further this project by taking these two factors into consideration. Lastly, simultaneous local field potential recordings may also be a necessary extension in order to reveal population coding deficits in the SynGAP$^{+/-}$ animals, as recent work has shown that *Syngap* gene mutation leads to differences in gamma range oscillations during exploratory and locomotive behaviour [134, 135].

In terms of models, it would be of great interest to further explore the recent developments in RNN modelling of neural population activity [130, 131]. While interpretation here still remains a primary concern, work has very recently been undertaken to reverse engineer these 'black box' models, applying sentiment classification techniques in order to extract interpretable descriptions of how they model particular systems [136]. This leads into the wider drive to combine systems neuroscience with other aspects of deep learning. While the two fields have always been closely linked, for example, through the fact that artificial neural networks take inspiration from biological network architectures in the brain, and in turn have displayed features of retinal mechanisms in object recognition tasks [137], researchers have recently called for a more unified approach to tackling neuroscience problems, stating that advancements in deep learning can greatly benefit the theoretical understanding and experimental progress of systems neuroscience [138].

6.3 Conclusions

To conclude, a novel manipulation to the Ising model was presented and explored in order to analyse the correlation structure of simulated and recorded networks as they changed over time. Results offered a successful proof of concept and opened doors to potential new and interesting forms of analysis in neural populations. Additionally, work using the diagonal Fisher information matrix revealed a highly informative and computationally efficient method for characterising large scale network structure and quantifying differences in this structure over time. Furthermore, due to the similarities between these approximated diagonals and the fully computed matrices at the smaller network level ($N = 10 - 20$), analysis presented in this thesis has opened the door to new exciting possibilities of large-scale network characterisation.

In relation to quantifying coding deficits in a mouse model of ASD, it was discovered that no significant differences between SynGAP$^{+/-}$ and WT genotypes were found in the population wide measures of imaged, awake-behaving mice in response to the prescribed experimental protocol. While this finding was somewhat surprising, it is highly interesting and furthers the knowledge of how behavioural phenotypes of intellectual disability manifest themselves at the level of neural population mechanisms.

Appendix A

A.1 Population Activity

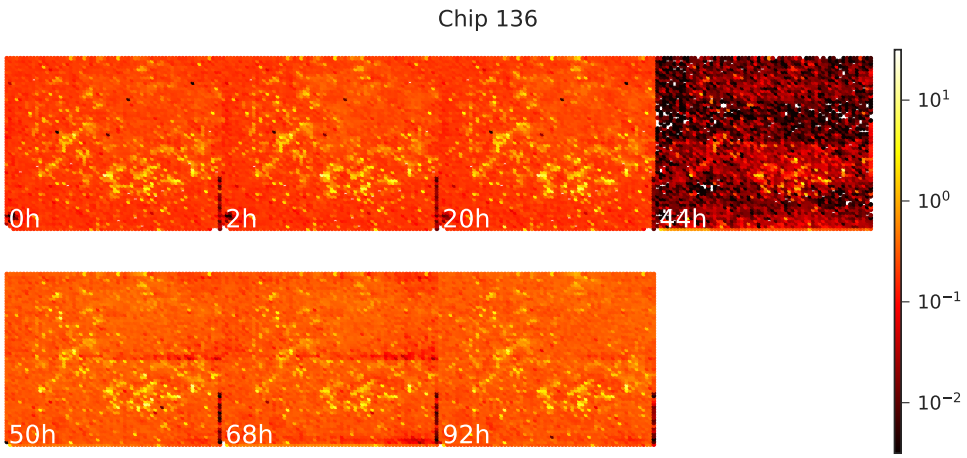


Figure A.1: Spatial activity at different time points throughout the experiment

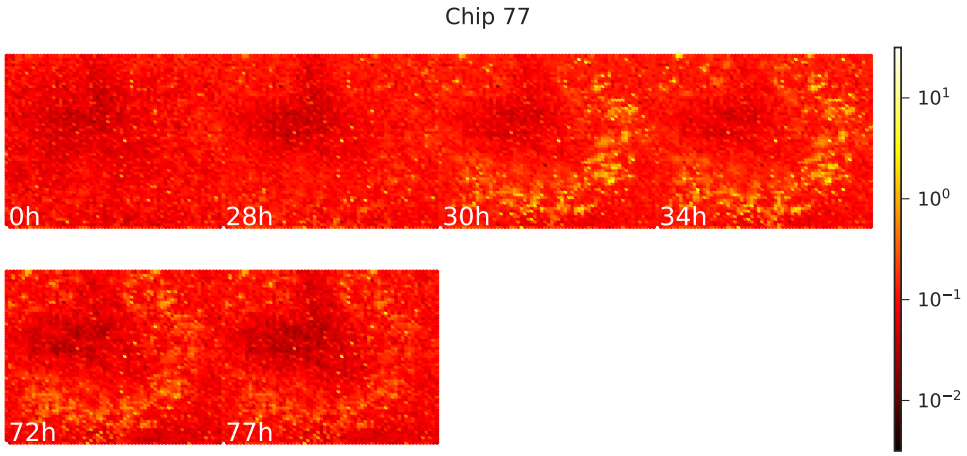


Figure A.2: Spatial activity at different time points throughout the experiment

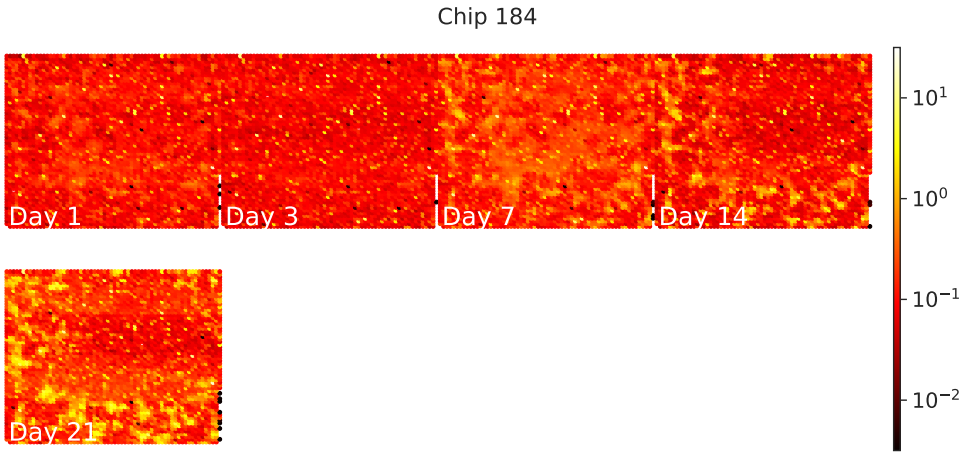


Figure A.3: Spatial activity at different time points throughout the experiment

Bibliography

[1] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, 1952.

[2] T. Toni and M. P. H. Stumpf. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26:104–110, 2010.

[3] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.

[4] G. Tkacik and W. Bialek. *Cell Biology: Networks, Regulation and Pathways*. Springer, 2007.

[5] T. O'Leary, A. C. Sutton, and E. Marder. Computational models in the age of large datasets. *Current Opinion in Neurobiology*, 32:87–94, 2015.

[6] P. S. Churchland and T. J. Sejnowski. *The computational brain*. Cambridge, MA: MIT Press, 1994.

[7] J. Humplik and G. Tkacik. Probabilistic models for neural populations that naturally capture global coupling and criticality. *PLOS Computational Biology*, 13(9):e1005763, 2017.

[8] G. Tkacik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek. Thermodynamics and signatures of criticality in a network of neurons. *Proceedings of the National Academy of Sciences*, 112(37):11508–11513, 2015.

[9] M. Nonnenmacher, C. Behrens, P. Berens, M. Bethge, and J. H. Macke. Signatures of criticality arise from random subsampling in simple population models. *PLoS Computational Biology*, 13(10):e1005718, 2017.

[10] T. Morra, S. Deny, and O. Marre. Dynamical criticality in the collective activity of a population of retinal neurons. *Physical Review Letters*, 114(7):078105, 2015.

[11] G. Hahn, A. Ponce-Alvarez, C. Monier, G. Benvenuti, A. Kumar, F. Chavance, G. Deco, and Y. Fregnac. Spontaneous cortical activity is transiently poised close to criticality. *PLoS Computational Biology*, 13(5):1–29, 2017.

[12] J. K. Miller, I. Ayzenshtat, L. Carrillo-Reid, and R. Yuste. Visual stimuli recruit intrinsically generated cortical ensembles. *PNAS*, 111(38):4053–4061, 2014.

[13] P. J. Uhlaas, G. Pipa, B. Lima, L. Melloni, S. Neuenschwander, D. Nikoliç, and W. Singer. Neural synchrony in cortical networks: history, concept and current status. *Frontiers Integrative Neuroscience*, 3:17, 2009.

[14] G. Buzsáki. Neural syntax: cell assemblies, synapsembles and readers. *Neuron*, 68(3):362–385, 2010.

[15] M. Okun, N. A. Steinmetz, L. Cossell, M. F. Iacaruso, H. Ko, P. Bartho, T. Moore, S. B. Hofer, T. D. Mrsic-Flogel, M. Carandini, and K. D. Harris. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521:511–515, 2015.

[16] M. Okun, P. Yger, S. L. Marguet, F. Gerard-Mercier, A. Benucci, S. Katzner, L. Busse, M. Carandini, and K. D. Harris. Population rate dynamics and multi-neuron firing patterns in sensory cortex. *Journal of Neuroscience*, 32(48), pages = 17108-17119), 2012.

[17] B. B. Averbeck, P. E. Latham, and A. Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7:358–366, 2006.

[18] J. Shlens, G. D. Field, J. L Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilinsky. The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, 26(32):8254–8266, 2006.

[19] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states. *Nature*, 440, 2006.

[20] G. Tkacik, T. Mora, O. Marre, D. Amodei, M. J. Berry, and W. Bialek. Searching for collective behavior in a large network of sensory neurons. *PLOS Computational Biology*, 10:e1003408, 2014.

[21] P. Peretto. Collective properties of neural networks: A statistical physics approach. *Biological Cybernetics*, 50:51–62, 1984.

[22] G. Tkacik, T. Mora, O. Marre, D. Amodei, M. J. Berry, and W. Bialek. Thermodynamics of networks of neurons: Signatures of criticality. *q-bio.NC*, arXiv:1407.5946, 2014.

[23] G. Tkacik, E. Schneidmann, M. J. Berry, and W. Bialek. Thermodynamics of networks of neurons: Signatures of criticality. *q-bio.NC*, arXiv:0611072, 2006.

[24] G. Tkacik E. Schneidmann, M. J. Berry, and W. Bialek. Spin glass models for a network of real neurons. *q-bio.NC*, arXiv:0912.5409, 2009.

[25] G. Tkacik, T. Mora, O. Marre, D. Amodei, M. J. Berry, and W. Bialek. The simplest maximum entropy model for collective behavior in a neural network. *q-bio.NC*, arXiv:1207.6319, 2012.

[26] A. Tang, D. Jackson, J. Hobbs, W. Chen, and J. L. Smith. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience*, 28(2):505–518, 2008.

[27] M. Okatan, M. A. Wilson, and E. N. Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation*, 17(9):1927–1961, 2005.

[28] E. Ising. *Contributions to the theory of Ferromagnetism*. PhD thesis, Institute of Physics, Hamburg, 1925.

[29] J. A. Hertz, Y. Roudi, and J. Trycha. Ising models for inferring network structure from spike data. *Quantitative Biology*, arXiv:1106.1752, 2011.

[30] J. A. Hertz, Y. Roudi, and J. Trycha. The ising model for neural data: Model quality and approximate methods for extracting functional connectivity. *Quantitative Biology*, arXiv:0902.2885, 2009.

[31] G. Tavoni, S. Cocco, and R. Monasson. Neural assemblies revealed by inferred connectivity-based models of prefontal cortex recordings. *Journal of Computational Neuroscience*, 41(3):269–293, 2016.

[32] C. O'Donnell, J. T. Gonçlaves, N. Whiteley, C. Portera-Cailliau, and T. J. Sejnowski. The population tracking model: A simple, scalable statistical model for neural population data. *Neural Computation*, 29:50–93, 2017.

[33] C. O'Donnell, J. T. Gonçlaves, C. Portera-Cailliau, and T. J. Sejnowski. Beyond excitation/inhibition imbalance in multidimensional models of neural circuit changes in brain disorders. *Elife*, 6:e26724, 2017.

[34] S. Cocco, R. Monasson, L. Posani, and G. Tavoni. Functional networks from inverse modelling of neural population activity. *Current Opinion in Systems Biology*, 3:103–110, 2017.

[35] C. Savin and G. Tkacik. Maximum entropy models as a tool for building precise neural controls. *Current Opinion in Systems Biology*, 46:120–126, 2017.

[36] C. Gardella, O. Marre, and T. Mora. Modelling the correlated activity of neural populations: a review. *Neural Computation*, 31(2):233–269, 2019.

[37] C. Gardella, O. Marre, and T. Mora. A trackable method for describing complex couplings between neurons and population rate. *eNeuro*, 3(4):0160–15, 2016.

[38] U. Koster, J. Sohl-Dickstein, C. M. Gray, and B. A. Olshausen. Modelling higher-order correlations within cortical microcolumns. *PLOS Compulational Biology*, 10(7), 2014.

[39] D. Spicher. Modelling multi-neuron spike trains with energy based models. Master's thesis, Institute of Adaptive and Neural Computation, The University of Edinburgh, 2014.

[40] J. T. Cronin. Statistical modelling and dimensionality reduction of simulated neural spike trains. Master's thesis, School of Informatics, The University of Edinburgh, 2015.

[41] D. Panas, H. Amin, A. Maccione, O. Muthmann, M. van Rossum, L. Berdondini, and M. H. Hennig. Sloppiness in spontaneously active neuronal networks. *J Neurosci*, 35:8480 – 8492, 2015.

[42] A. Ly, M. Marsman, J. Verhagen, R. Grasman, and E. Wagenmakers. A tutorial on fisher information. *Journal of Mathematical Psychology*, 80:40–55, 2017.

[43] J. Martens. New insights and perspectives on the natural gradient method. *arXiv*, page 1412.1193, 2014.

[44] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell. Overcoming catastrophic forgetting in neural networks. *PNAS*, 114:3521–3526, 2017.

[45] I. Kanitscheider, R. Coen-Cagli, A. Kohn, and A. Pouget. Measuring fisher information accurately in correlated neural populations. *PLoS Comp Biol*, 12(6):e1004218, 2015.

[46] A. Ponce-Alvarez, G. Mochol, J. de la Rocha, and G. Deco. Cortical state transition and stimulus response evolve along stiff and sloppy parameter directions, respectively. *bioRxiv*, 2019.

[47] G. Buzsaki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7:446–451, 2004.

[48] F. Nault and P. De Koninck. *Dissociated Hippocampal Cultures. In: Doering L. (eds) Protocols for Neural Cell Culture*. Springer Protocols Handbooks. Humana Press, 2009.

[49] M. Corner. Spontaneous neuronal burst dishcharges as dependent and independent variables in the maturation of cerebral cortex tissue cultures *in vitro*: A review of activity-dependent studies in live 'model' systems for the development of intrinsically generated bioelectric slow-wave sleep patterns. *Brain Re Rev*, 59:221–244, 2008.

[50] P. Berkes, G. Orban, M. Lengyel, and J. Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331:83–87, 2011.

[51] J. Fiser, M. Lengyel, P. Berkes, and G. Orban. How (not) to assess the importance of correlations for the matching of spontaneous and evoked activity. *Quantitative Biology*, arXiv:1301.6554, 2013.

[52] I. H. Stevenson and K. P. Kording. How advances in neural recording affect data analysis. *Nature Neuroscience*, 14:139–142, 2011.

[53] M. L. Andermann, A. M. Kerlin, D. K. Roumis, L. L. Glickfeld, and R. C. Reid. Functional specialization of mouse higher visual cortical areas. *Neuron*, 72(6):1025–1039, 2010.

[54] J. W. Gibbs. *Thermodynamics*. Longmans, Green and Company, 1906.

[55] E. Ising. A contribution to the theory of ferromagnetism. *Phys*, 31(1):253–258, 1925.

[56] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 2012.

[57] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Ann. Math. Statist.*, 43(5):1470–1480, 1972.

[58] C. R. Rao. Information and accuracy attainable in estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37:81–91, 1945.

[59] S. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society, 2007.

[60] B. B. Mavhta, R. Chachra, M. K. Transtrum, and J. P. Sethna. Parameter space compression underlies emergent theories and predictive models. *Science*, 342:604–607, 2013.

[61] G. E. Crooks. Measuring thermodynamic length. *Phys Rev Letters*, 99:100602–6, 2007.

[62] J. P. Sethna. What are sloppy models? http://www.lassp.cornell.edu/sethna/Sloppy/WhatAreSloppyModels.html, 2008. Online; accessed 24-January-2016.

[63] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3:e189, 2007.

[64] K. S. Brown and J. P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review*, 68:021904, 2003.

[65] K. S. Brown and J. P. Sethna. Sloppiness, robustness, and evolvability in systems biology. *Opinion Biotech*, 19:389 – 395, 2008.

[66] R. Pascanu and Y. Bengio. Revisiting natural gradient for deep networks. *arXiv*, page 1301.3584, 2013.

[67] A. Gonzalez and J. Dorronsoro. Natural conjugate gradient training of multilayer perceptions. *Artificial Neural Networks ICANN*, pages 169–177, 2006.

[68] D. C. Dowson and B. V. Landau. The frechet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12(3):450–455, 1982.

[69] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

[70] T. M. Cover and J. A. Thomas. Elements of information theory. *Wiley-Interscience*, 2, 2006.

[71] K. H. Boven, M. Fejtl, A. Moller, W. Nisch, and A. Stett. Advances in network electrophysiology using multi-electrode arrays. *New York: Springer Press*, pages 24–37, 2006.

[72] M. E. J. Obien, K. Deligkaris, T. Bullmann, and D. J. Bakkum. Revealing neuronal function through microelectrode array recordings. *Frontiers in Neuroscience*, 8:423, 2015.

[73] J. O. Muthmann, H. Amin, E. Sernagor, A. Maccione, D. Panas, L. Berdondini, U. S. Bhalla, and M. H. Hennig. Spike detection for large neural populations using high density multielectrode arrays. *Frontiers in Neuroinformatics*, 9:28, 2015.

[74] A. Minerbi, R. Kahana, L. Goldfelf, M. Kaufman, S. Marom, and N. Ziv. Long-term relationships between synaptic tenacity, synaptic remodelling and network activity. *PLOS Biology*, 7:1–20, 2009.

[75] E. Slomowitz, B. Styr, I. Vertkin, H. Milshtein-Parush, I. Nelken, M. Slutsky, and I. Slutsky. Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLIFE*, 4:e04378, 2015.

[76] E. Cohen, M. Ivenshitz, V. Amor-Baroukh, V. Greenberger, and M. Segal. Determinants of spontaneous activity in networks of cultured hippocampus. *Brain Research*, 1235:21–30, 2008.

[77] J. G. Orlandi, J. Soriano, E. Alvarez-Lacalle, S. Teller, and J. Casademunt. Noise focusing and the emergence of coherent activity in neuronal cultures. *Nature Physics*, 9:582–590, 2013.

[78] Y. Penn, M. Segal, and E. Moses. Network synchronization in hippocampal neurons. *PNAS*, 113:3341–3346, 2016.

[79] P. J. Attwell, S. Rahman, M. Ivarsson, and C. H. Yeo. Cerebellar AMPA-kainate receptor blockade prevents performance of classically conditioned nictitating mebrane responses. *Journal of Neuroscience*, 19(24):RC45, 1999.

[80] B. Cokić and V. Stein. Stargazin modulated AMPA receptor antagonism. *Neuropharmacology*, 58(7):1062–1070, 2008.

[81] Q. Li and B. D. Burrell. CNQX and AMPA inhibit electrical synaptic transmission: a potential interaction between electrical and glutamatergic synapses. *Brain Res.*, 1228:43–57, 2008.

[82] T. L. Carter and J. G. McElligott. Cerebellar AMPA/KA receptor antagonism by CNQX inhibits vestibuloocular reflex adaptation. *Experimental Brain Research*, 166(2):157–169, 2005.

[83] J. O. Muthmann. *Multiscale Modelling of homeostatic processes*. PhD thesis, Institute of Adaptive and Neural Computation, School of Informatics, The University of Edinburgh, 2017.

[84] L. Tian, S. A. Hires, and L. L. Looger. Imaging neuronal activity with genetically encoded calcium indicators. *Cold Spring Harbor Protocols*, 7(6):647–656, 2012.

[85] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79(8):2554–2558, 1982.

[86] J. H. Macke, P. Berens, A. S. Ecker, A. S. Tolias, and M. Bethge. Generating spike trains with specified correlation coefficients. *Neural Computation*, 21:397–423, 2009.

[87] J. H. Macke, M. Opper, and M. Bethge. Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Physical Review Letters*, 106:208102, 2011.

[88] J. H. Macke, M. Opper, and M. Bethge. The effect of pairwise neural correlations on global population statistics, 2011.

[89] S. Yu, H. Yang, H. Nakahara, G. S. Santos, D. Nikolic, and D. Plenz. Higher-order interactions characterized in cortical activity. *Journal of Neuroscience*, 31(48):17514–17526, 2011.

[90] G. Buzsaki and K. Mizuseki. The log-dynamic brain: how skewed distributions affect network operations. *Nature reviews Neuroscience*, 15(4):264–278, 2014.

[91] A. Roxin, N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk. On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience*, 31(45):16217–16226, 2011.

[92] T. Hromadka and M. R. Deweese. Sparse representation of sounds in the unanesthetized auditory cortex. *PLos Biology*, 6(1):e16, 2008.

[93] P. C. Petersen and R. W. Berg. Lognormal firing rate distribution reveals prominent fluctuationdriven regime in spinal motor networks. *eLife*, 5:e18805, 2016.

[94] L. Takács. On the method of inclusion and exclusion. *J. Amer. Statist. Assoc.*, 62:102–113, 1967.

[95] A. Onken, S. Grünewälder, M. Munk, and K. Obermayer. Modelling short-term noise dependence of spike counts in macaque prefrontal cortex. *Advances in Neural Information Processing Systems*, 21, 2008.

[96] D. Panas. *Model-based analysis of stability in networks of neurons*. PhD thesis, Institute of Adaptive and Neural Computation, School of Informatics, The University of Edinburgh, 2017.

[97] R. Perou, R. H. Bitsko amd S. J. Blumberg, P. Pastor, S. Avenevoli, M. D. Kogan, and L. N. Huang. Mental health surveillance among children - United States, 2005-2011. *MMWR Suppl.*, 62:1–35, 2013.

[98] A. K. Srivastava and C. E. Schwartz. Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neuroscience and biobehavioral reviews*, 46:161–174, 2014.

[99] H. N. Komoyama, A. M. Watabe, H. J. Carlisle, K. Porter, S. J. Martin, R. G. Morris, T. J. O'Dell, and S. G. Grant. SynGAP regulates ERK/MAPK signalling,synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J Neuroscience*, 22:9721–9732, 2002.

[100] J. P. Clement, M. Aceti, T. K. Creson, E. D. Ozkan, Y. Shi, N. J. Reish, A. G. Almonte, B. H. Miller, B. J. Wiltgen, C. A. Miller, X. Xu, and G. Rumbaugh. Pathogenic SYNGAP1 mutations impair cognitive development by disrupting maturation of dendritic spine synapses. *Cell*, 151:709–723, 2012.

[101] N. Jeyabalan and J. P. Clement. SYNGAP1: Mind the gap. *Frontiers in Cellular Neuroscience*, 10:32, 2016.

[102] A. D. Holtmaat, J. T. Trachtenberg, L. Wilbrecht, G. M. Shepard, X. Zhang, G. W. Knott, and K. Svoboda. Transient and persistent dendritic spines in the neocortex in vivo. *Neuron*, 45:279–291, 2005.

[103] L. E. Vazquez, H. J. Chen, I. Sokolova, I. Knuesel, and M. B. Kennedy. SynGAP regulates steady-state and activity dependent phosphorylation of cofilin. *J Neuroscience*, 28:8862–8872, 2008.

[104] H. J. Carlisle, P. Manzerra, E. Marcora, and M. B. Kennedy. SynGAP regulates spine formation. *J Neuroscience*, 24:13673–13683, 2004.

[105] T. Rose, J. Jaepel, M. Hubener, and T. Bonhoeffer. Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science*, 352(6291):1319–1322, 2016.

[106] M. Weldon, M. Kilinc, J. L. Holder, and G. Rumbaugh. The first international conference on SYNGAP1-related brain disorders: a stakeholder meeting of families, researchers, clinicians, and regulators. *Journal of Neurodevelopmental Disorders*, 10:6, 2018.

[107] S. W. Keemink, S. C. Lowe, J. M. P. Pakan, E. Dylda, M. C. W. van Rossum, and N. L. Rochefort. FISSA: A neuropil decontamination toolbox for calcium imaging signals. *Scientific Reports*, 8(1):3493, 2018.

[108] T. Deneux, A. Kaszas, G. Szalay an G. Katona, T. Lakner, A. Grinvald, B. Rozsa, and I. Van Zetta. Accurate spike estimation from noisy calcium sig-

nals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nature*, 7:12190, 2016.

[109] M. Kaneko, Y. Fu, and M. P. Stryker. Locomotion induces stimulus-specific response enhancement in adult visual cortex. *Journal of Neuroscience*, 37(13):3532–3543, 2017.

[110] A. Ranson. Stability and plasticity of contextual modulation in the mouse visual cortex. *Cell Reports*, 18(4):1–31, 2016.

[111] K. Ohki, S. Chung, Y. H. Ch'ng, P. Kara, and R. C. Reid. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, 433:597–603, 2005.

[112] M. Mazurek, M. Kager, and S. D. Van Hooser. Robust quantification of orientation selectivity and direction selectivity. *Frontiers in Neural Circuits*, 8:92, 2014.

[113] B. Bhaumik and N. P. Shah. Development and matchin of binocular orientation preference in mouse V1. *Frontiers in Systems Neuroscience*, 8:128, 2014.

[114] C. X. He, D. A. Cantu, S. S. Mantri, W. A. Zeiger, A. Goel, and C. Portera-Cailliau. Tactile defensiveness and impaired adaptation of neuronal activity in the Fmr1 knock-out mouse model of autism. *Journal of Neuroscience*, 37(27):6475–6487, 2017.

[115] J. T. Gonçalves, J. E. Anstey, P. Golshani, and C. Portera-Cailliau. Circuit level defects in the developing neocortex of fragile X mice. *Nat Neuroscience*, 16(7):903–909, 2013.

[116] S. A. Barnes, L. S. Wijetunge, A. D. Jackson, D. Katsanevaki, E. K. Osterweil, N. H. Komiyama, S. G. Grant, M. F. Bear, U. V. Nagerl, P. C. Kind, and D. J. Wyllie. Convergence of hippocampal pathophysiology in Syngap+/- and Fmr1-/y mice. *Journal of Neuroscience*, 35(45):15073–15081, 2015.

[117] E. Dylda. *Neuronal Circuits of Experience-Dependent Plasticity in the Primary Visual Cortex*. PhD thesis, Centre for Integrative Physiology, School of Biomedical Sciences, The University of Edinburgh, 2018.

[118] S. R. Schultz and S. Panzeri. Temporal correlations and neural spike train entropy. *Physical Review Letters*, 86(25):5823–5826, 2001.

[119] A. Dettner, S. Münzberg, and T. Tchumatchenko. Temporal pairwise spike correlations fully capture single-neuron information. *Nature Communications*, 7(13805), 2016.

[120] T. Tchumatchenko, T. Geisel, M. Volgushev, and F. Wolf. Spike correlations - what can they tell about synchrony? *Frontiers in Neuroscience*, 5:68, 2011.

[121] G. Hinton and A. Brown. Spiking Boltzmann machines. In *Proceedings of the 13th Conference on Advances in Neural Informaion Processing Systems, NIPS*, 2000.

[122] I. Sutskever, G. Hinton, and G. W. Taylor. The recurrent restricted Boltzmann machine. In *Proceedings of the 21st Conference on Advances in Neural Informaion Processing Systems, NIPS*, 2008.

[123] J. C. Angles d'Auriac and R. Rammal. On kinetic Ising models in one dimension. *Journal of Physics A: Mathematical and Theoretical*, 21:763–768, 1988.

[124] A. Witoelar and Y. Roudi. Neural network reconstruction using kinetic Ising models with memory. *BMC Neuroscience*, 12(Suppl 1):274, 2011.

[125] A. I. Weber and J. W. Pillow. Capturing the dynamical repertoire of single neurons with generalized linear models. *arXiv*, page 1602.07389, 2016.

[126] J. Shlens. Notes on generalized linear models of neurons. *arXiv*, page 1404.1999, 2014.

[127] S. Gerwinn, J. H. Macke, and M. Bethge. Bayesian inference for generalized linear models for spiking neurons. *Frontiers in Computational Neuroscience*, 4:12, 2010.

[128] V. Lawhern, W. Wu, N. Hatsopoulos, and L. Paninski. Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of Neuroscience Methods*, 189(2):267–280, 2010.

[129] R. Kobayashi, S. Kurita, A. Kurth, K. Kitano, K. Mizuseki, M. Diesmann, B. J. Richmond, and S. Shinomoto. Reconstructing neuronal circuitry from parallel spike trains. *Nature Communications*, 10(1):4468, 2019.

[130] N. Maheswaranathan, A. H. Williams, M. D. Golub, S. Ganguli, and D. Sussilo. Universality and individuality in neural dynamics across large populations of recurrent networks. *arXiv*, page 1907.08549, 2019.

[131] A. Nayebi, D. Bear, J. Kubilius, K. Kar, S. Ganguliand D. Sussillo, J. J. DiCarlo, and D. L. Yamins. Task-driven convolutional recurrent models of the visual system. In *Advances in Neural Information Processing Systems 31*, pages 5290–5301. Curran Associates, Inc., 2018.

[132] E. Lee, J. Lee, and E. Kim. Excitation/inhibition imbalance in animal models of autism spectrum disorders. *Biological Psychiatry*, 81(10):838–847, 2017.

[133] R. Gao and P. Penzes. Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Current Molecular Medicine*, 15(2):146–167, 2015.

[134] M. H. Berryer, B. Chattopadhyaya, P. Xing, I. Riebe, C. Bosoi, N. Sanon, J. Antoine-Bertrand, M. Léveque, M. Avoli, F. F. Hamdan, L. Carmant, N. Lamarche-Vane, J. C. Lacaille, J. L. Michaud, and G. Di Cristo. Decrease of SYNGAP1 in GABAergic cells impairs inhibitory synapse sonnectivity, synaptic inhibition and cognitive function. *Nature Communications*, 9(7):13340, 2016.

[135] M. Vinck, R. Batista-Brito, U. Knoblich, and J. A. Cardin. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 86(3):740–754, 2015.

[136] N. Maheswaranathan, A. H. Williams, M. D. Golub, S. Ganguli, and D. Sussilo. Reverse engineering recurrent networks for sentiment classification reveals line attractor dynamics. *arXiv*, page 1906.10720, 2019.

[137] L. T. McIntosh, N. Maheswaranathan, A. Nayebi, S. Ganguli, and S. A. Baccus. Deep learning models of the retinal response to natural scenes. *arXiv*, page 1702.01825, 2017.

[138] B. A. Richards, T. P. Lillicrap, and P. Beaudoin. A deep learning framework for neuroscience. *Nature Neuroscience*, 22:1761–1770, 2019.